

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: June 10, 2003, 22:00:04 ; Search time 8532.25 Seconds

(without alignments)
10921.753 Million cell updates/sec

Title: US-09-502-426b-1_COPY_1_3202
Perfect score: 3202
Sequence: 1 atgtgggtatctatctgtg.....gagagagagaactagctcc 3202

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 2054640 segs, 14551402878 residues

Total number of hits satisfying chosen parameters: 4109280

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%

Listing first 45 summaries

Database :

GenEmbl.*
1: gb_ba:*
2: gb_htg:*
3: gb_in:*
4: gb_om:*
5: gb_ov:*
6: gb_pat:*
7: gb_ph:*
8: gb_pl:*
9: gb_pr:*
10: gb_ro:*
11: gb_sy:*
12: gb_un:*
13: gb_vl:*
14: gb_vl:*
15: em_ba:*
16: em_fun:*
17: em_hum:*
18: em_in:*
19: em_mu:*
20: em_om:*
21: em_or:*
22: em_ov:*
23: em_pat:*
24: em_ph:*
25: em_pl:*
26: em_ro:*
27: em_sts:*
28: em_un:*
29: em_vl:*
30: em_htg_hum:*
31: em_htg_inv:*
32: em_htg_other:*
33: em_htg_mus:*
34: em_htg_pln:*
35: em_htg_rod:*
36: em_htg_mam:*
37: em_htg_vrt:*
38: em_sy:*
39: em_htgo_hum:*
40: em_htgo_mus:*
41: em_htgo_other:*

score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
C 1	3134.2	97.9	84196	8 ATT3A5	AL132979 Arabidops
2	1132	35.4	4818	11 AF044216	AF044216 Arabidops
3	176.8	5.5	191	11 AL844732	AL844732 Arabidops
4	133.2	4.2	104992	2 AC005504	AC005504 Plasmodu
5	133.2	4.2	169546	2 AC004157	AC004157 Plasmodu
C 6	130.8	4.1	104992	2 AC005504	AC005504 Plasmodu
7	130.8	4.1	169546	2 AC004157	AC004157 Plasmodu
C 8	129.4	4.0	14867	3 AE001398	AE001398 Plasmodu
9	128.6	4.0	172816	9 AC093899	AC093899 Homo sapi
C 10	127.8	4.0	86827	3 PPMAL3P5	AL034556 Plasmodu
C 11	126.4	3.9	133877	2 AC120883	AC120883 Homo sapi
C 12	125.4	3.9	86827	3 PPMAL3P5	AL034556 Plasmodu
C 13	124.6	3.9	205130	2 AC105425	AC105425 Homo sapi
14	124.4	3.9	4601	3 DMU11584	U11584 Drosophila
15	124.4	3.9	19517	3 DMU37541	U137541 Drosophila
C 16	123.4	3.9	67970	3 PPMAL1P3	AL031746 Plasmodu
C 17	122.6	3.8	4601	3 DMU11584	U11584 Drosophila
C 18	122.6	3.8	19517	3 DMU37541	U137541 Drosophila
C 19	121.2	3.8	177293	2 AC069525	AC069525 Homo sapi
C 20	121	3.8	172816	9 AC093899	AC093899 Homo sapi
21	121	3.8	176670	9 AC104073	AC104073 Homo sapi
22	118	3.7	67970	3 PPMAL1P3	AL031746 Plasmodu
23	118	3.7	84196	8 ATT3A5	AL132979 Arabidops
C 24	116.4	3.6	72243	9 AL731858	AL731858 Human DNA
25	115.6	3.6	205130	2 AC105425	AC105425 Homo sapi
26	115.2	3.6	133877	2 AC120883	AC120883 Homo sapi
C 27	114.4	3.6	176670	9 AC104073	AC104073 Homo sapi
C 28	113	3.5	14867	3 AE001398	AE001398 Plasmodu
C 29	112.6	3.5	137889	9 AC073269	AC073269 Homo sapi
C 30	112.2	3.5	206059	2 AC127383	AC127383 Homo sapi
31	111.6	3.5	104853	9 AC117444	AC117444 Homo sapi
32	111.4	3.5	111861	9 AC069435	AC069435 Homo sapi
C 33	110.2	3.4	15656	3 AB070264	AB070264 Bombyx mo
C 34	110.2	3.4	15656	3 AB070264	AB070264 Bombyx mo
35	109.6	3.4	159618	9 AC0696750	AC0696750 Homo sapi
C 36	107.8	3.4	152209	9 HS1108D11	AL034419 Human DNA
37	107	3.3	105238	9 AC011458	AC011458 Homo sapi
38	106.6	3.3	106958	9 AL807813	AL807813 Human DNA
39	106.4	3.3	155106	9 AC104069	AC104069 Homo sapi
40	106.2	3.3	181792	9 AC098822	AC098822 Homo sapi
41	105.6	3.3	176174	9 AC007483	AC007483 Homo sapi
C 42	105.6	3.3	177293	2 AC069525	AC069525 Homo sapi
43	105.2	3.3	206059	2 AC127383	AC127383 Homo sapi
44	104.8	3.3	114276	9 AC011355	AC011355 Homo sapi
45	104.6	3.3	127439	9 AC126178	AC126178 Homo sapi

ALIGNMENTS

RESULT 1
ATT3A5 84196 bp DNA linear PLN 26-JAN-2000
LOCUS Arabidopsis thaliana DNA chromosome 3, BAC clone T3A5.
DEFINITION Arabidopsis thaliana
ACCESSION AL132979
VERSION AL132979.2 GI:6782244
KEYWORDS
SOURCE
ORGANISM Arabidopsis thaliana.
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
Rosidae; eurosids II; Brassicales; Brassicaceae; Arabidopsids.
REFERENCE
1 (bases 1 to 84196)
Bloeker, H., Mewes, H.W., Lemcke, K., Mayer, K.F.X., Quetler, F. and
Salanoubat, M.

JOURNAL Unpublished
REFERENCE 2 (bases 1 to 84196)
AUTHORS EU Arabidopsis sequencing project.
TITLE Direct Submission
JOURNAL Submitted (25-JAN-2000) MIPS, at the Max-Planck-Institut fuer Biochemie, Am Klopferspitz 18a, D-82152 Martinsried, FRG, E-mail: lemcx@mips.biochem.mpg.de, mayer@mips.biochem.mpg.de, Project Coordinator: Marcel Salanoubat and Francis Queller, Groupement d'Interet Public, Centre National de Sequence - GENOSCOPE, 2 rue Gaston Cremieux, BP191, 91006 Evry Cedex, France;
http://www.genoscope.cns.fr
COMMENT On Jan 27, 2000 this sequence version replaced gi:6434247. Information on performance of analysis and a more detailed annotation of this entry and other sequences of chromosomes 3, 4 and 5 can be viewed at: <http://www.mips.biochem.mpg.de/proj/thal/>.

FEATURES
source
1..84196
/organism="Arabidopsis thaliana"
/variety="Columbia"
/db_xref="taxon:3702"
/chromosome="3"
2983..3338
/note="355 bp LTR"
3339..7564
/note="LTR-transposon"
7565..7922
/note="367bp LTR"
12339..13197
/gene="T3A5.10"
join(12339..12665,12742..12803,12888..13055,13125..13197)
/gene="T3A5.10"
/note="similarity to cyclin-dependent kinase inhibitor (ICK1), Arabidopsis thaliana, EMBL:AT094772_1"
/codon_start=1
/product="putative protein"
/protein_id="CAB62432.1"
/db_xref="GI:6561966"
/translation="MAAVRRREDDVYBENCVTITTKRRKMEBEVDLVESRITLPCV QATNRGIVARNASGASSETSVIVARRRDSPEVEQCIIEEDSSVSCSTSEKSKRR IEFVLDLENNDDRETETSWIYDILNKSESMNMDSSVAVEDVESRRIRKSLHETV KEAELEDFQVAKDLRNLKLLCCSMKYNFDEKDEPLGGGYEMVKLP"
12339..12665
/gene="T3A5.10"
/number=1
12666..12741
/gene="T3A5.10"
/number=1
12742..12803
/gene="T3A5.10"
/number=2
12804..12887
/gene="T3A5.10"
/number=2
12888..13055
/gene="T3A5.10"
/number=3
13056..13124
/gene="T3A5.10"
/number=3
13125..13197
/gene="T3A5.10"
/number=4
13198..16470
/gene="T3A5.20"
complement(join(15858..15901,16014..16470))
/note="similarity to predicted proteins, Arabidopsis thaliana"
/codon_start=1
/product="putative protein"
/protein_id="CAB62433.1"
/db_xref="GI:6561967"
/translation="MARLEMHNSMGCAARISFSEFEVIRSEKSNKSNINRSRSFS MSADAFSVTDSIMIPADEIFLKGKILPKKETS SHVHRLGEEELLTEEGSMDGNTF

SLRPILSSSSFSSTKGTWRELLGKRTYHRSKKTDKVNEEVLSDQHKITSGNATRECC QVADTR"
complement(15858..15901)
/number=1
complement(15902..16013)
/number=1
complement(16014..16470)
/number=2
18303..19931
/gene="T3A5.30"
complement(18303..19931)
/codon_start=1
/product="scarecrow-like 7 (SCL7)"
/protein_id="CAB62434.1"
/db_xref="GI:6561968"
/translation="MAYMCTDSNLMAIAOOLIKOKOOOOSOHOOEOEORPEPNPNS PSFGFTLPGSGFSDPQVNTDPCGHFTHHONAAVASSEFSDMESELLINGDAS QTPNDPRTIYGHDPFVSFPRSLSPSYINRNRKDSASQOLPPPPASIAIYSPSPSQ HPPPPPPDPDINQPTFKAIHDYARKPEPTDPTLRIKESVSESGDITQVGTFAE ALSHKETESPSSSSSLEDFTILSYKTLNDACYSKFAHLTAQALILEATVNSNNIHI VDFGIFGQIOWSALQALATRSSGKPTRISGIPAPSLDSDGSPSLIATGNRLRDA AILDNEEFYPLVPLIOLNNGSSFRVDQVLYVNFMLELKLDELATVGTALRLA RSLNPRTVLGEYVSLNRYEPANRYKNSLRFSVAFESLEPNDLRDSKRLRYERYL FGRRINDLVSDDDNNKPGTRFLMEKEEDMRVLMERAGFEPPYKPSNYAVSOAKLLIM NYNISTYLSLVESEPGFISLAKNNVPLTLTSSMR"
complement(18303..19931)
/number=1
26093..28999
/gene="T3A5.40"
complement(join(26093..26338,27091..27200,27308..27386, 27478..27570,27723..28037,28136..28288,28374..28698, 28779..28999))
/note="Contains Cytochrome P450 cysteine heme-1ron ligand signature AA45-464;Prokaryotic membrane lipoprotein lipid attachment site AA320-330"
/codon_start=1
/product="steroid 22-alpha-hydroxylase (DWF4)"
/protein_id="CAB62435.1"
/db_xref="GI:6561969"
/translation="MFEETHTHTLPLPLPSLSLLEFLILKRRNRKTRFNPCKS GMPFLGFTIGYLKPYATTLTQMOQHSYKGIKRSNLGEPPIYASADGLNRFTIQ NEGRLFECSYPRSIGIGITGLKSMYLIVGDMHMRSTSLNPLSHARLRTILLDVERH TLPVLDSMOONSIFSADDEAKKFTFNMAKHISMNPGSEETSQLKEYUTPEKGYVS APJNLPEGTATYHKAQSRATILKFTERMEERKLDIKEDDEEVEVKEDAEBSKSDH VRKQRDDDLGLWVLRKHSNLSTEQIIDLILSLFAGETSSVAIALAIFPLQACPAKV EELREBELLETARAKKELGESELMWDDYKKKDFQCVINETLRGLNVVRFJHRAKADY RYKGYDIPSGMKVLPVISAVALDINSRDQNLNPNMDOOQNNGASSSGSGFTWGN NYMPFGGGRPLCAGSELAKLEMAVFTHHLVLKKNWLEADDKCFAPFVDFPGLPIR VSRIL"
complement(26093..26338)
/number=1
complement(26339..27090)
/number=1
complement(27091..27200)
/number=2
complement(27201..27307)
/number=2
complement(27308..27386)
/number=3
complement(27387..27477)
/number=3
complement(27478..27570)
/number=4
complement(27571..27722)
/number=4
complement(27723..28037)
/number=5
complement(28038..28135)
/number=5
complement(28136..28288)
/number=6
complement(28289..28373)
/number=6

QY	421	AGTTTATTTGATCATATATACATGATCATTTCTTGATTACTATGAAGATTTCTTAC	480
Db	31780	AGGTTTATTTGATCATATATATACATGATCATTTCTTGATTACTATGAAGATTTCTTAC	31721
QY	481	CAATTAAATTTGGAATTCATCTCTGTGATTATTAATTAATTAATGAGTGAATATCC	540
Db	31720	CAATTAAATTTGGAATTCATCTCTGTGATTATTAATTAATTAATGAGTGAATATCC	31661
QY	541	GTTATATCGATCTCCATTCATGATATGATTTCTTGCTTAATCCAGCAATATTATACA	600
Db	31660	GTTATATCGATCTCCATTCATGATATGATTTCTTGCTTAATCCAGCAATATTATACA	31601
QY	601	AGAGTATTGAAAAAAACGAAATATAGAAAAAGGAAAGATGATGCCATGAGATG	660
Db	31600	AGAGTATTGAAAAAAACGAAATATAGAAAAAGGAAAGATGATGCCATGAGATG	31541
QY	661	TGATATATTTCAAAAGGATATAGATATGACAAACAAAGGTGTGGATATATGGTCCCT	720
Db	31540	TGATATATTTCAAAAGGATATAGATATGACAAACAAAGGTGTGGATATATGGTCCCT	31481
QY	721	GCCAGCTTTTCTCACATCATATGACCCCTATTTGATTTTCTGGATATATCGTTAA	780
Db	31480	GCCAGCTTTTCTCTCACATCATATGACCCCTATTTGATTTTCTGGATATATCGTTAA	31421
QY	781	TTTGCATTAACGATTTGTGAAAAAATTTTATTTTGTAGCTGATCTCAATATATGTCCA	840
Db	31420	TTTGCATTAACGATTTGTGAAAAAATTTTATTTTGTAGCTGATCTCAATATATGTCCA	31361
QY	841	GGTATTTTGCAATATCTTCTGTTTAAAGCATATTTGTCTTCTTTTGTGTTGCTTCT	900
Db	31360	GGTATTTTGCAATATCTTCTGTTTAAAGCATATTTGTCTTCTTTTGTGTTGCTTCT	31301
QY	901	TAACTATATATATCGGGATATATGATATATCAATGATATATACAAACAAATATGCTGG	960
Db	31300	TAACTATATATATCGGGATATATGATATATCAATGATATATACAAACAAATATGCTGG	31241
QY	961	ACCAATTTTGAATTAACCTTTTCTCAAAATTTACGGGACATGAGCTCGACCTTAAATA	1020
Db	31240	ACCAATTTTGAATTAACCTTTTCTCAAAATTTACGGGACATGAGCTCGACCTTAAATA	31182
QY	1021	CGATTTTACAGCGCTACTAGTATGATTTACTACATTAAGCATTAAGAGCCGTTACAGC	1080
Db	31181	CGATTTTACAGCGCTACTAGTATGATTTACTACATTAAGCATTAAGAGCCGTTACAGC	31122
QY	1081	TATTTATTAACAAGTTACAACTGAAATTTACTGGAATCCCTTAGAAAAATTTTGGAAAT	1140
Db	31121	TATTTATTAACAAGTTACAACTGAAATTTACTGGAATCCCTTAGAAAAATTTTGGAAAT	31062
QY	1141	CCGGTGTATATGAATATATAGATTTACTGTGTAACAAATATGTTATCAATTAAGTGCTA	1200
Db	31061	CCGGTGTATATGAATATATAGATTTACTGTGTAACAAATATGTTATCAATTAAGTGCTA	31002
QY	1201	ACATATACATATATTCCTTACAGAAAAACAACTTAGAGAGATTAACATATCATATAT	1260
Db	31001	ACATATACATATATTCCTTACAGAAAAACAACTTAGAGAGATTAACATATCATATAT	30942
QY	1261	GGGATATGCTATACCTTTCACGTATGATCTATCTAGAGACTTAAGCAATATGTTATGTATGTC	1320
Db	30941	GGGATATGCTATACCTTTCACGTATGATCTATCTAGAGACTTAAGCAATATGTTATGTATGTC	30882
QY	1321	GATAAATGAATTCACACGCGTGTATTAATTTATGGAGCCGATGTTACGATCACTGCA	1380
Db	30881	GATAAATGAATTCACACGCGTGTATTAATTTATGGAGCCGATGTTACGATCACTGCA	30822
QY	1381	ATATCATTTCTTGTTGGTCAACAATTAATAACAAACACAGAAAAAGAAAAAGATTTT	1440
Db	30821	ATATCATTTCTTGTTGGTCAACAATTAATAACAAACACAGAAAAAGAAAAAGATTTT	30762
QY	1441	CTGGATTCACATTAATGATCTTAATAGCATATCTTTGGGTTTCAATTTGSAAGTCC	1500
Db	30761	CTGGATTCACATTAATGATCTTAATAGCATATCTTTGGGTTTCAATTTGSAAGTCC	30702

			/strain="Arabidopsis thaliana"				
			/organelle="ns-2"				
			/db_xref="taxon:3702"				
			/chromosome="III"				
			/map="8 cm from ngac"				
			990.. 4398				
			/gene="DMF4"				
			/note="dwarf4"				
			/allele="dwf4-1; dwf4-2; dwf4-3; dwf4-4"				
			990.. 1055				
			/gene="DMF4"				
			join(<1133.. 1353,1434.. 1158,1844.. 1996,2095.. 2409,				
			2562.. 2654,2746.. 2824,2931.. 3040,3795.. 4398)				
			/gene="DMF4"				
			/product="steroid 22-alpha-hydroxylase"				
			join(1133.. 1353,1434.. 1158,1844.. 1996,2095.. 2409,				
			2562.. 2654,2746.. 2824,2931.. 3040,3795.. 4040)				
			/gene="DMF4"				
			/function="brassinosteroid biosynthetic enzyme"				
			/note="member of the cytochrome P450 superfamily; CYP90B1"				
			/codon_start=1				
			/product="steroid 22-alpha-hydroxylase"				
			/protein_id="AAC05093.1"				
			/db_xref="GI:2935342"				
			/translation="MEFEETHTLLEPLLSLLFLILKRRNRKTRNIPGKS				
			GMPLEATIGIKPYTATTTGLDEPMQWVSVYGRYLNLGEPTVSADGNINPQL				
			NEGLPEFCSPRSISQADEAKFTFNMAKHSLMDGEEETDLKEEYTEMKSDH				
			TLRYLDMSWONSIFSAODEAKFTFNMAKHSLMDGEEETDLKEEYTEMKSDH				
			APLPLPGTAHKALOSRAITLKLERMERKLDIKEDEEEVEKVEDAEKSGDV				
			VKRQRDDDLGLGWLIKHSNSTEQIDILLIILSFAGHETSVAIALAIFLOACPAA				
			EELHEELAIARAKEGESELMNDMDYKKDPOCVINETLRGVNVPFLHRKLKVQ				
			RKYQDIPISGWKVLPIVASVHLDSRYDNPFLPMPMWOONNGASSGGSSFTGMEN				
			NYMFGGSRPLCAGSELAKLEMAYFIHHLYLKFWEIAEDDPFAFPFVDPNLPRLR				
			VSRLL"				
BASE COUNT	1566 a	727 c	888 g	1637 t			
ORIGIN							
Query Match		35.4%	Score 1132;	DB 8; Length 4818;			
Best Local Similarity		100.0%;	Pred. No. 1.1e-153;				
Matches 1132; Conservative	0;	Mismatches	0;	Indels	0;	Gaps	0;
OY	2071	GAATTCAAAATTCCTCATTAATAATTTTTTGAAATCTACAAATTAATAATTTAGTCAAT	2130				
Db	1	GAATTCAAAATTCCTCATTAATAATTTTTTGAAATCTACAAATTAATAATTTAGTCAAT	60				
OY	2131	AACAATGCATGGAAGAAGTCCAAAAAAAATTTGTGTAACAGAAACTGCCAATTTTTTTT	2190				
Db	61	AACAATGCATGGAAGAAGTCCAAAAAAAATTTGTGTAACAGAAACTGCCAATTTTTTTT	120				
OY	2191	TTTATGGAACAAGAAATTAACAGATAGAAACTAATTTGTTGGAAAGGAATAGTATATA	2250				
Db	121	TTTATGGAACAAGAAATTAACAGATAGAAACTAATTTGTTGGAAAGGAATAGTATATA	180				
OY	2251	TACATTAAAGCAAATTTAAAAAATTTATATAAGCCTATACGGCGTCAAAAGTAGTATCTA	2310				
Db	181	TACATTAAAGCAAATTTAAAAAATTTATATAAGCCTATACGGCGTCAAAAGTAGTATCTA	240				
OY	2311	GTAGTGTAAATTAATAATGATGGTGGCATTCACAAATTTGGACAAACAATGAAAAAGCAAT	2370				
Db	241	GTAGTGTAAATTAATAATGATGGTGGCATTCACAAATTTGGACAAACAATGAAAAAGCAAT	300				
OY	2371	TAAATATTAATCTTAAAAATTAATAAAAAATTTGAGTAAATGTGTTTCTGACTATTGAGG	2430				
Db	301	TAAATATTAATCTTAAAAATTAATAAAAAATTTGAGTAAATGTGTTTCTGACTATTGAGG	360				
OY	2431	GGCAAAAAAAAAACAAATGCCAAAAGTCTACGGGTTTGACTGTCCAGTTCCGATTAATCT	2490				
Db	361	GGCAAAAAAAAAACAAATGCCAAAAGTCTACGGGTTTGACTGTCCAGTTCCGATTAATCT	420				
OY	2491	AATAACCTGTGCTTTGACCGACGCTGTGTAGGGGTCCTTCACATTTTCACGTCTCT	2550				
Db	421	AATAACCTGTGCTTTGACCGACGCTGTGTAGGGGTCCTTCACATTTTCACGTCTCT	480				

QY	2551	ACCCCTACTGCTGAGCCCA	ACCCCTTTTCCCAATCTCTAAGGGTAATTTTGGAAATCCCAAT	2610	
Db	481	ACCCCTACTGCTGAGCCCA <th>ACCCCTTTTCCCAATCTCTAAGGGTAATTTTGGAAATCCCAAT</th> <th>540</th>	ACCCCTTTTCCCAATCTCTAAGGGTAATTTTGGAAATCCCAAT	540	
OY	2611	TTAAACGATTTGACACCGTAC <th>CGGACTTCGGGATTCGCTGAGCATTTATCAAAAT</th> <th>2670</th>	CGGACTTCGGGATTCGCTGAGCATTTATCAAAAT	2670	
Db	541	TTAAACGATTTGACACCGTAC <th>CGGACTTCGGGATTCGCTGAGCATTTATCAAAAT</th> <th>600</th>	CGGACTTCGGGATTCGCTGAGCATTTATCAAAAT	600	
OY	2671	TATTAGCAGCAATGGGTTTAT	TATTTAAAAACTCACAACTTGATCAGATAAAAATTCAT	2730	
Db	601	TATTAGCAGCAATGGGTTTAT	TATTTAAAAACTCACAACTTGATCAGATAAAAATTCAT	660	
OY	2731	AAACACTTTTACGTGATTCGT <th>ACGATCTATCTAAAGACTTTTTTTTCTACACGGT</th> <th>2790</th>	ACGATCTATCTAAAGACTTTTTTTTCTACACGGT	2790	
Db	661	AAACACTTTTACGTGATTCGT <th>ACGATCTATCTAAAGACTTTTTTTTCTACACGGT</th> <th>720</th>	ACGATCTATCTAAAGACTTTTTTTTCTACACGGT	720	
OY	2791	GGATGAAGATATATGTACTAT	TATTCACGAGCAATTTGATTTATGATATATCAATTAATCC	2850	
Db	721	GGATGAAGATATATGTACTAT	TATTCACGAGCAATTTGATTTATGATATATCAATTAATCC	780	
OY	2851	ATGATATTTATGATATATAAT	AGCTTAAACTATTTTCAGATCCAGCTTTCTGCACACTT	2910	
Db	781	ATGATATTTATGATATATAAT	AGCTTAAACTATTTTCAGATCCAGCTTTCTGCACACTT	840	
OY	2911	TTGTTTTTAAATTTAAGACTT	TAAATATAAAGATTTAAAGAGCATATACGAGGCACAA	2970	
Db	841	TTGTTTTTAAATTTAAGACTT	TAAATATAAAGATTTAAAGAGCATATACGAGGCACAA	900	
OY	2971	AAAGTAATGAACACGGAGAA <th>CAAAAGCCATGAAGCTCATGTTAGTTTAAGCTTAATA</th> <th>3030</th>	CAAAAGCCATGAAGCTCATGTTAGTTTAAGCTTAATA	3030	
Db	901	AAAGTAATGAACACGGAGAA <th>CAAAAGCCATGAAGCTCATGTTAGTTTAAGCTTAATA</th> <th>960</th>	CAAAAGCCATGAAGCTCATGTTAGTTTAAGCTTAATA	960	
OY	3031	AGAAGATTTTATTAATTTAAT	TATATGACATGATTAACAATTAATTTTTCGACTTCTTTAAA	3090	
Db	961	AGAAGATTTTATTAATTTAAT	TATATGACATGATTAACAATTAATTTTTCGACTTCTTTAAA	1020	
OY	3091	ACCCCTCTTTCACAAACGAC <th>AGCTCCCTTTTTCAGTAGACAGTCCGATTCGCCATCTTAAAG</th> <th>3155</th>	AGCTCCCTTTTTCAGTAGACAGTCCGATTCGCCATCTTAAAG	3155	
Db	1021	ACCCCTCTTTCACAAACGAC <th>AGCTCCCTTTTTCAGTAGAGTCCGATTCGCCATCTTAAAG</th> <th>1080</th>	AGCTCCCTTTTTCAGTAGAGTCCGATTCGCCATCTTAAAG	1080	
OY	3151	ACAAAGCATTAGAAAGAAAG <th>AGTAGAGAGAGAGAGAGAGAAACTACTCTCC</th> <th>3202</th>	AGTAGAGAGAGAGAGAGAGAAACTACTCTCC	3202	
Db	1081	ACAAAGCATTAGAAAGAAAG <th>AGTAGAGAGAGAGAGAGAGAAACTACTCTCC</th> <th>1132</th>	AGTAGAGAGAGAGAGAGAGAAACTACTCTCC	1132	
RESULT 3	AL844732	191 bp	DNA	linear	STS 25-JUL-2002
LOCUS	DEFINITION	Arabidopsis thaliana transposon insertion	STS SM.3.29533	sequence	tagged site.
ACCESSION	AL844732	1	GI:21998461		
VERSION	AL844732.1				
KEYWORDS	STS: STS	sequence tagged site.			
SOURCE	ORGANISM	Arabidopsis thaliana.			
REFERENCE	AUTHORS	Clarke, J.H., Bowles, B., Carter, J., Hart, D., McCullagh, B., Murphy, G., Langham, S., LeGrys, C., Jones, J.D.G. and Bevan, M.			
JOURNAL	REFERENCE	Unpublished			
AUTHORS	TITLE	2 (bases 1 to 191)			
JOURNAL	COMMENT	Clarke, J.H.			
		Submitted (25-JUL-2002)			
		Lane, Norwich, NK4 7UJ, UK			
		AT denotes an activation tag dissociation transposon within a			
		single line, ET an enhancer trap dissociation transposon, GT a			
		trap dissociation transposon, MT a mis-expression enhancer trap			
		dissociation transposon, SM a defective suppressor mutator			
		transposon, _3 denotes a sequence derived from the 3' end of the			
		transposon, _5 denotes a sequence derived from the 5' end of the			

OY	699	AGCTTGGAATATGTCCTCCGACGCTTCTCTCAATCATATGACCTATTTGG	758
Db	85724	ATATATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT	85665
OY	759	ATTTCTGATATTCGTTAAATTTGGCATACGATTTGCAAAATTTTATTTGTTAG	818
Db	85664	TAATTTAAATTAATTAATTAAGA-----AACATATTTTAACTTTTA	85623
OY	819	CTGATCTCAATATTTATGTTCCAGTATTTGCAATATCTTCTGTTAAAGCATATTTGTC	878
Db	85622	TTTAAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT	85564
OY	879	TTTCTTTTGTTCGTTCTCTTAACATATATTTACGGGATATATGATACATATGTA	938
Db	85563	TTTATTTTATTTATTAATTAATTAATTAATTTTAAATATATTTTAAATTAATTAAT	85504
OY	939	TATCACAAAACATTTGTCGGACCATTTTGAATTAACCTTTTCTCAAAACATTTCCGGAC	998
Db	85503	ATATTAACATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT	85444
OY	999	ACTGACCTGACCTTAATTAATTCATTTTACAGGCTCATGTTGATGATACATACATA	1058
Db	85443	ATATTTACATTTTATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT	85384
OY	1059	AGCATTAAGGACCGCTTCAACCTTTTATACAAAGTCAAACTGAATATAGCTGAAT	1118
Db	85383	AAATTAATTAATTAATTTGCTGAATTAATTAATTAATTAATTAATTAATTAAT	85324
OY	1119	CTTTAGAAATTTTGGATTTACCGGTTGATATGTAATTAATTAATTAATTAATTAAT	1178
Db	85323	TATATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT	85264
OY	1179	TA-----TGTATCAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT	1234
Db	85263	AAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT	1294
OY	1235	TAAAGAAAGTTAATCAATATATATGATATGATATGATATGATATGATATGAT	1348
Db	85203	TTATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT	1408
OY	1295	AGCTTAAGAA-----TATGATATGATATGATATGATATGATATGATATGAT	1468
Db	85143	ATATATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT	1528
OY	1349	AAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT	1584
Db	85083	TTATTTTATTTATTTATTTATTTATTTATTTATTTATTTATTTATTTATTTAT	1608
OY	1409	AACAAACACAGAAAGAAAGAAAGATTTTCTGATGATTCATTCATGATTAATG	1668
Db	85023	ATTTATTTATTTATTTATTTATTTATTTATTTATTTATTTATTTATTTATTTAT	1728
OY	1469	CATGATCTTTTGGGTTACAGTTTCGAGACGCTCAAGCGTTACATGCACTTA	1784
Db	84963	TTTATTAATTAATTTATTTATTTATTTATTTATTTATTTATTTATTTATTTAT	1848
OY	1529	TTAAATTCCTTTTAAAGCATTTTAAACATATTTATTTATTTATTTATTTATTTAT	1908
Db	84903	ATATTAATTTATTTATTTATTTATTTATTTATTTATTTATTTATTTATTTAT	1968
OY	1589	GCGAATCTGTAACATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT	2028
Db	84843	TATATATTTATTTATTTATTTATTTATTTATTTATTTATTTATTTATTTAT	2088
OY	1649	TACTTGGATGTTAACTTAATCTTGTTCCTCAAGTATTAATTAATTAATTAATTAAT	2148
Db	84783	TATTAATTAATTAATTTATTTATTTATTTATTTATTTATTTATTTATTTATTTAT	2208
OY	1709	GTAAGAAAGTTTGTCTTATTTTCGATGATGATGATGATGATGATGATGATGAT	2268
Db	84723	ATATTTATTTATTTATTTATTTATTTATTTATTTATTTATTTATTTATTTAT	2328
OY	1769	TTTGGAAATGTAACCTTTTACATGATTAATTAATTAATTAATTAATTAATTAAT	2388
Db	84664	TTTGGAAATGTAACCTTTTACATGATTAATTAATTAATTAATTAATTAATTAAT	2448
OY	1828	TTTGGAAATGTAACCTTTTACATGATTAATTAATTAATTAATTAATTAATTAAT	2508
Db	84663	ATTAATTAATTTCTTTAATTTATTAATTAATTAATTTATTTATTTATTTATTTAT	2568
OY	1829	GACAGCCTTCACACATGATGATGATGATGATGATGATGATGATGATGATGAT	2628
Db	84603	TTTAAATTAATTTATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT	2688
OY	1889	GCTATGATCAATGAAAGAAAGCCTGAGATTAATTAATTAATTAATTAATTAAT	2748
Db	84548	ATTAATTAATTTATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT	2808
OY	1949	AAAAATGTAATTAATTTTCAACATTTTATTAATTAATTAATTAATTAATTAAT	2868
Db	84488	ATTAATTAATTTATTTATTAATTAATTAATTAATTAATTAATTAATTAATTAAT	2928
OY	2009	GTTCTTAAATTTTATTTATTTATTTATTTATTTATTTATTTATTTATTTATTTAT	2988
Db	84428	TTAATTAATTAATTTATTTATTTATTTATTTATTTATTTATTTATTTATTTAT	3048
OY	2069	GTAATTAATTAATTTATTTATTTATTTATTTATTTATTTATTTATTTATTTAT	3108
Db	84369	TAAATTAATTAATTTATTTATTTATTTATTTATTTATTTATTTATTTATTTAT	3168
OY	2129	ATTAACATTAATTAATTTATTTATTTATTTATTTATTTATTTATTTATTTATTTAT	3228
Db	84309	ATTAATTAATTTATTTATTTATTTATTTATTTATTTATTTATTTATTTATTTAT	3288
OY	2189	TTTATTAATTAATTTATTTATTTATTTATTTATTTATTTATTTATTTATTTAT	3348
Db	84249	ATTAATTAATTTATTTATTTATTTATTTATTTATTTATTTATTTATTTATTTAT	3408
OY	2249	TATACATTAAGCAATTTATTTATTTATTTATTTATTTATTTATTTATTTATTTAT	3468
Db	84194	TATATTAATTTATTTATTTATTTATTTATTTATTTATTTATTTATTTATTTAT	3528
OY	2280	ATTAATTAATTTATTTATTTATTTATTTATTTATTTATTTATTTATTTATTTAT	3588
Db	84163	ATTAATTAATTTATTTATTTATTTATTTATTTATTTATTTATTTATTTATTTAT	3648
OY	2280	ATTAATTAATTTATTTATTTATTTATTTATTTATTTATTTATTTATTTATTTAT	3708
Db	84195	ATTAATTAATTTATTTATTTATTTATTTATTTATTTATTTATTTATTTATTTAT	3768
OY	2280	ATTAATTAATTTATTTATTTATTTATTTATTTATTTATTTATTTATTTATTTAT	3828
Db	84195	ATTAATTAATTTATTTATTTATTTATTTATTTATTTATTTATTTATTTATTTAT	3888
OY	2280	ATTAATTAATTTATTTATTTATTTATTTATTTATTTATTTATTTATTTATTTAT	3948
Db	84195	ATTAATTAATTTATTTATTTATTTATTTATTTATTTATTTATTTATTTATTTAT	4008
OY	2280	ATTAATTAATTTATTTATTTATTTATTTATTTATTTATTTATTTATTTATTTAT	4068
Db	84195	ATTAATTAATTTATTTATTTATTTATTTATTTATTTATTTATTTATTTATTTAT	4128
OY	2280	ATTAATTAATTTATTTATTTATTTATTTATTTATTTATTTATTTATTTATTTAT	4188
Db	84195	ATTAATTAATTTATTTATTTATTTATTTATTTATTTATTTATTTATTTATTTAT	4248
OY	2280	ATTAATTAATTTATTTATTTATTTATTTATTTATTTATTTATTTATTTATTTAT	4308
Db	84195	ATTAATTAATTTATTTATTTATTTATTTATTTATTTATTTATTTATTTATTTAT	4368
OY	2280	ATTAATTAATTTATTTATTTATTTATTTATTTATTTATTTATTTATTTATTTAT	4428
Db	84195	ATTAATTAATTTATTTATTTATTTATTTATTTATTTATTTATTTATTTATTTAT	4488
OY	2280	ATTAATTAATTTATTTATTTATTTATTTATTTATTTATTTATTTATTTATTTAT	4548
Db	84195	ATTAATTAATTTATTTATTTATTTATTTATTTATTTATTTATTTATTTATTTAT	4608
OY	2280	ATTAATTAATTTATTTATTTATTTATTTATTTATTTATTTATTTATTTATTTAT	4668
Db	84195	ATTAATTAATTTATTTATTTATTTATTTATTTATTTATTTATTTATTTATTTAT	4728
OY	2280	ATTAATTAATTTATTTATTTATTTATTTATTTATTTATTTATTTATTTATTTAT	4788
Db	84195	ATTAATTAATTTATTTATTTATTTATTTATTTATTTATTTATTTATTTATTTAT	4848
OY	2280	ATTAATTAATTTATTTATTTATTTATTTATTTATTTATTTATTTATTTATTTAT	4908
Db	84195	ATTAATTAATTTATTTATTTATTTATTTATTTATTTATTTATTTATTTATTTAT	4968
OY	2280	ATTAATTAATTTATTTATTTATTTATTTATTTATTTATTTATTTATTTATTTAT	5028
Db	84195	ATTAATTAATTTATTTATTTATTTATTTATTTATTTATTTATTTATTTATTTAT	5088
OY	2280	ATTAATTAATTTATTTATTTATTTATTTATTTATTTATTTATTTATTTATTTAT	5148
Db	84195	ATTAATTAATTTATTTATTTATTTATTTATTTATTTATTTATTTATTTATTTAT	5208
OY	2280	ATTAATTAATTTATTTATTTATTTATTTATTTATTTATTTATTTATTTATTTAT	5268
Db	84195	ATTAATTAATTTATTTATTTATTTATTTATTTATTTATTTATTTATTTATTTAT	5328
OY	2280	ATTAATTAATTTATTTATTTATTTATTTATTTATTTATTTATTTATTTATTTAT	5388
Db	84195	ATTAATTAATTTATTTATTTATTTATTTATTTATTTATTTATTTATTTATTTAT	5448
OY	2280	ATTAATTAATTTATTTATTTATTTATTTATTTATTTATTTATTTATTTATTTAT	5508
Db	84195	ATTAATTAATTTATTTATTTATTTATTTATTTATTTATTTATTTATTTATTTAT	5568
OY	2280	ATTAATTAATTTATTTATTTATTTATTTATTTATTTATTTATTTATTTATTTAT	5628
Db	84195	ATTAATTAATTTATTTATTTATTTATTTATTTATTTATTTATTTATTTATTTAT	5688
OY	2280	ATTAATTAATTTATTTATTTATTTATTTATTTATTTATTTATTTATTTATTTAT	5748
Db	84195	ATTAATTAATTTATTTATTTATTTATTTATTTATTTATTTATTTATTTATTTAT	5808
OY	2280	ATTAATTAATTTATTTATTTATTTATTTATTTATTTATTTATTTATTTATTTAT	5868
Db	84195	ATTAATTAATTTATTTATTTATTTATTTATTTATTTATTTATTTATTTATTTAT	5928
OY	2280	ATTAATTAATTTATTTATTTATTTATTTATTTATTTATTTATTTATTTATTTAT	5988
Db	84195	ATTAATTAATTTATTTATTTATTTATTTATTTATTTATTTATTTATTTATTTAT	6048
OY	2280	ATTAATTAATTTATTTATTTATTTATTTATTTATTTATTTATTTATTTATTTAT	6108
Db	84195	ATTAATTAATTTATTTATTTATTTATTTATTTATTTATTTATTTATTTATTTAT	6168
OY	2280	ATTAATTAATTTATTTATTTATTTATTTATTTATTTATTTATTTATTTATTTAT	6228
Db	84195	ATTAATTAATTTATTTATTTATTTATTTATTTATTTATTTATTTATTTATTTAT	6288
OY	2280	ATTAATTAATTTATTTATTTATTTATTTATTTATTTATTTATTTATTTATTTAT	6348
Db	84195	ATTAATTAATTTATTTATTTATTTATTTATTTATTTATTTATTTATTTATTTAT	6408
OY	2280	ATTAATTAATTTATTTATTTATTTATTTATTTATTTATTTATTTATTTATTTAT	6468
Db	84195	ATTAATTAATTTATTTATTTATTTATTTATTTATTTATTTATTTATTTATTTAT	6528
OY	2280	ATTAATTAATTTATTTATTTATTTATTTATTTATTTATTTATTTATTTATTTAT	6588
Db	84195	ATTAATTAATTTATTTATTTATTTATTTATTTATTTATTTATTTATTTATTTAT	6648
OY	2280	ATTAATTAATTTATTTATTTATTTATTTATTTATTTATTTATTTATTTATTTAT	6708
Db	84195	ATTAATTAATTTATTTATTTATTTATTTATTTATTTATTTATTTATTTATTTAT	6768
OY	2280	ATTAATTAATTTATTTATTTATTTATTTATTTATTTATTTATTTATTTATTTAT	6828
Db	84195	ATTAATTAATTTATTTATTTATTTATTTATTTATTTATTTATTTATTTATTTAT	6888
OY	2280	ATTAATTAATTTATTTATTTATTTATTTATTTATTTATTTATTTATTTATTTAT	6948
Db	84195	ATTAATTAATTTATTTATTTATTTATTTATTTATTTATTTATTTATTTATTTAT	7008
OY	2280	ATTAATTAATTTATTTATTTATTTATTTATTTATTTATTTATTTATTTATTTAT	7068
Db	84195	ATTAATTAATTTATTTATTTATTTATTTATTTATTTATTTATTTATTTATTTAT	7128
OY	2280	ATTAATTAATTTATTTATTTATTTATTTATTTATTTATTTATTTATTTATTTAT	7188
Db	84195	ATTAATTAATTTATTTATTTATTTATTTATTTATTTATTTATTTATTTATTTAT	7248
OY	2280	ATTAATTAATTTATTTATTTATTTATTTATTTATTTATTTATTTATTTATTTAT	7308
Db	84195	ATTAATTAATTTATTTATTTATTTATTTATTTATTTATTTATTTATTTATTTAT	7368
OY	2280	ATTAATTAATTTATTTATTTATTTATTTATTTATTTATTTATTTATTTATTTAT	7428
Db	84195	ATTAATTAATTTATTTATTTATTTATTTATTTATTTATTTATTTATTTATTTAT	7488
OY	2280	ATTAATTAATTTATTTATTTATTTATTTATTTATTTATTTATTTATTTATTTAT	7548
Db	84195	ATTAATTAATTTATTTATTTATTTATTTATTTATTTATTTATTTATTTATTTAT	7608
OY	2280	ATTAATTAATTTATTTATTTATTTATTTATTTATTTATTTATTTATTTATTTAT	7668
Db	84195	ATTAATTAATTTATTTATTTATTTATTTATTTATTTATTTATTTATTTATTTAT	7728
OY	2280	ATTAATTAATTTATTTATTTATTTATTTATTTATTTATTTATTTATTTATTTAT	7788
Db	84195	ATTAATTAATTTATTTATTTATTTATTTATTTATTTATTTATTTATTTATTTAT	7848
OY	2280	ATTAATTAATTTATTTATTTATTTATTTATTTATTTATTTATTTATTTATTTAT	7908
Db	84195	ATTAATTAATTTATTTATTTATTTATTTATTTATTTATTTATTTATTTATTTAT	7968
OY	2280	ATTAATTAATTTATTTATTTATTTATTTATTTATTTATTTATTTATTTATTTAT	8028
Db	84195	ATTAATTAATTTATTTATTTATTTATTTATTTATTTATTTATTTATTTATTTAT	8088
OY	2280	ATTAATTAATTTATTTATTTATTTATTTATTTATTTATTTATTTATTTATTTAT	8148
Db	84195	ATTAATTAATTTATTTATTTATTTATTTATTTATTTATTTATTTATTTATTTAT	8208
OY	2280	ATTAATTAATTTATTTATTTATTTATTTATTTATTTATTTATTTATTTATTTAT	8268
Db	84195	ATTAATTAATTTATTTATTTATTTATTTATTTATTTATTTATTTATTTATTTAT	8328
OY	2280	ATTAATTAATTTATTTATTTATTTATTTATTTATTTATTTATTTATTTATTTAT	8388
Db	84195	ATTAATTAATTTATTTATTTATTTATTTATTTATTTATTTATTTATTTATTTAT	8448
OY	2280	ATTAATTAATTTATTTATTTATTTATTTATTTATTTATTTATTTATTTATTTAT	8508
Db	84195	ATTAATTAATTTATTTATTTATTTATTTATTTATTTATTTATTTATTTATTTAT	8568
OY	2280	ATTAATTAATTTATTTATTTATTTATTTATTTATTTATTTATTTATTTATTTAT	8628
Db	84195	ATTAATTAATTTATTTATTTATTTATTTATTTATTTATTTATTTATTTATTTAT	8688
OY	2280	ATTAATTAATTTATTTATTTATTTATTTATTTATTTATTTATTTATTTATTTAT	8748
Db	84195	ATTAATTAATTTATTTATTTATTTATTTATTTATTTATTTATTTATTTATTTAT	8808
OY	2280	ATTAATTAATTTATTTATTTATTTATTTATTTATTTATTTATTTATTTATTTAT	8868
Db	84195	ATTAATTAATTTATTTATTTATTTATTTATTTATTTATTTATTTATTTATTTAT	8928
OY	2280	ATTAATTAATTTATTTATTTATTTATTTATTTATTTATTTATTTATTTATTTAT	8988
Db	84195	ATTAATTAATTTATTTATTTATTTATTTATTTATTTATTTATTTATTTATTTAT	9048
OY	2280	ATTAATTAATTTATTTATTTATTTATTTATTTATTTATTTATTTATTTATTTAT	9108
Db	84195	ATTAATTAATTTATTTATTTATTTATTTATTTATTTATTTATTTATTTATTTAT	9168
OY	2280	ATTAATTAATTTATTTATTTATTTATTTATTTATTTATTTATTTATTTATTTAT	9228
Db	84195	ATTAATTAATTTATTTATTTATTTATTTATTTATTTATTTATTTATTTATTTAT	9288
OY	2280	ATTAATTAATTTATTTATTTATTTATTTATTTATTTATTTATTTATTTATTTAT	9348
Db	84195	ATTAATTAATTTATTTATTTATTTATTTATTTATTTATTTATTTATTTATTTAT	9408
OY	2280	ATTAATTAATTTATTTATTTATTTATTTATTTATTTATTTATTTATTTATTTAT	9468
Db	84195	ATTAATTAATTTATTTATTTATTTATTTATTTATTTATTTATTTATTTATTTAT	9528
OY	2280	ATTAATTAATTTATTTATTTATTTATTTATTTATTTATTTATTTATTTATTTAT	9588
Db	84195	ATTAATTAATTTATTTATTTATTTATTTATTTATTTATTTATTTATTTATTTAT	9648
OY	2280	ATTAATTAATTTATTTATTTATTTATTTATTTATTTATTTATTTATTTATTTAT	9708
Db	84195	ATTAATTAATTTATTTATTTATTTATTTATTTATTTATTTATTTATTTATTTAT	9768
OY	2280	ATTAATTAATTTATTTATTTATTTATTTATTTATTTATTTATTTATTTATTTAT	9828
Db	84195	ATTAATTAATTTATTTATTTATTTATTTATTTATTTATTTATTTATTTATTTAT	9888
OY	2280	ATTAATTAATTTATTTATTTATTTATTTATTTATTTATTTATTTATTTATTTAT	9948
Db	84195	ATTAATTAATTTATTTATTTATTTATTTATTTATTTATTTATTTATTTATTTAT	10008

FEATURES

source

JOURNAL

Medical Center Drive, Rockville, MD 20814, USA

REFERENCE

Submitted (02-NOV-1998) The Institute for Genomic Research, 9712

AUTHORS

Submitted (02-NOV-1998) The Institute for Genomic Research, 9712

TITLE

Submitted (02-NOV-1998) The Institute for Genomic Research, 9712

PUBMED

Submitted (02-NOV-1998) The Institute for Genomic Research, 9712

REFERENCE

Submitted (02-NOV-1998) The Institute for Genomic Research, 9712

AUTHORS

Submitted (02-NOV-1998) The Institute for Genomic Research, 9712

TITLE

Submitted (02-NOV-1998) The Institute for Genomic Research, 9712

PUBMED

Submitted (02-NOV-1998) The Institute for Genomic Research, 9712

REFERENCE

Submitted (02-NOV-1998) The Institute for Genomic Research, 9712

AUTHORS

Submitted (02-NOV-1998) The Institute for Genomic Research, 9712

TITLE

Submitted (02-NOV-1998) The Institute for Genomic Research, 9712

PUBMED

Submitted (02-NOV-1998) The Institute for Genomic Research, 9712

REFERENCE

Submitted (02-NOV-1998) The Institute for Genomic Research, 9712

AUTHORS

Submitted (02-NOV-1998) The Institute for Genomic Research, 9712

TITLE

Submitted (02-NOV-1998) The Institute for Genomic Research, 9712

PUBMED

Submitted (02-NOV-1998) The Institute for Genomic Research, 9712

REFERENCE

Submitted (02-NOV-1998) The Institute for Genomic Research, 9712

AUTHORS

Submitted (02-NOV-1998) The Institute for Genomic Research, 9712

TITLE

	QY	1597	GTAACATCAAACTTTTATTTATTTGATAGTCGTGATAGCATTTCCAAATCAATCCTTTGG	1656
	Dd	8645	ATTATGATATAAATTTTATATATATTAATTAATTTATTTAGTATATAAAAATTTAAACAATAT	8704
	QY	1657	ATGTTTAAACTTAATCTTGTTTTCTCCACGGAATTAATATTTAATCATCGAGGTAAA	1716
	Dd	8705	ACATTATTAATTTACACATTAATTAATTTATTTATTTAATTAATTTTACATTTATTTAAT	8764
	QY	1717	AGTTTTCCTATTTTTCGCATGCAAGAGATAAACCCTAATGCTTAATTTTTCAAA	1776
	Dd	8765	TAAATTAATTTTATTTATTTATTTATTTATTTATTTAATTAATTTATTTATATATTAAT	8824
	QY	1777	ATGTAACCCCTTTTCTCATAGATTAATTTACGCTGTTGTTTTGTCGCATATGACACCT	1836
	Dd	8825	TTATTTTATTTTATTTTATTTATTT - TTATTTTATTTTATTTATTTATTTATTTTATTTT	8883
	QY	1837	CTACACTGTGATAGTCAAATTTTTCGCCAATATTAATTAATTAGAATTCAGTCTACTAT	1896
	Dd	8884	ATTTATTTAATTAATTTATTTAATTAATTTTAAATTAATTAATTTATTTATTTATTTAA	8943
	QY	1897	CAATAGAAGAACGCTGAGATATACATTTTAATTTAAGCAAAATTTTGA AAAATGT	1956
	Dd	8944	TTATTTTAAATTAATTTATTTATTTTATTTATTTATTTATTTAATTTAATTTATTTAA	9003
	QY	1957	TATATTTTCACAAATATTTATTAATAATATGATGCGCTTAATATGATTTCCATGCTCTAA	2016
	Dd	9004	TATATTTATTTATTTATTTATTTATTTATTTATTTATTTAATTAATTTATTTATTAATTA	9063
	QY	2017	AATATTTTTTTTATTTATTTAGTTAVAAAATACATTATGAACCAATATAGTTGG-----T	2070
	Dd	9064	ATTTATTTTAAATTAATTTATTTATTTATTTATTTTAAATTTATTTATTTATTTATTTAAAT	9123
	QY	2071	GAATTCAAATATCCCAATTAATTTTGTGAATCFCAAAATTTATTAATTTAGTCAT	2130
	Dd	9124	TATTTATTTTATTTATTTATTTATTTATTTATTTATTTATTTATTTATTTATTTATTTAT	9183
	QY	2131	AACAATGCATAGAAAGTCCAAAAAAATTTTGTAAACAGAACTGCCAAATTTTTTTTT	2190
	Dd	9184	TATTTTAGTTATTTATTTATTTATTTAATTAATTTATTTATTTATTTATTTATTTATTTAT	9243
	QY	2191	TTTATGCAACAGAAATPACAGATAGAAACTATTTTGTGTGGAATGGAATGAGATGATATA	2250
	Dd	9244	TTTTTAATTTTATTTATTTATTTAATTAATTTATTTAATTTAATTTTAAATTTATTTATTT	9303
	QY	2251	TACATTPAAGCAATTTTAAAAAATTTATATAAGCCTATRAGCGGCCAAAGTATGTAACCTA	2310
	Dd	9304	TAAATTAATTTATTTATTTATTTTATTTAATTAATTTATTTTAAATTAATTTATTTATATA	9363
	QY	2311	GTAGGTGAATTAATTAATGCAATGCTGCATTCAGAAATTTGGAGCAACAATGAAACGGAAAT	2370
	Dd	9364	TTAATTTTAAAAAATATTTTAAATTTATTT - TTATTTAATCTTATATTTTATATTTTAC	9421
	QY	2371	TAAAAATATTAACCTTAAAAATATATPAATAATTTGAGTAATAGTGTTCGACTATGAGG	2430
	Dd	9422	ACACATTTTATTTTATATAAAAAAATATATCTTTTAAATATATATATTAATTAATTAAT	9481
	QY	2431	GGCAAAAAAAGACATAGCCAAAAAGTCTA	2459
	Dd	9482	ACTAAAAATTTAAAAATATATATTAATGATATA	9510
	RESULT 9			
	LOCUS	AC093899	172816 bp	DNA linear PRI 12-JUN-2002
	DEFINITION	Homo sapiens BAC clone RP11-724016 from 2, complete sequence.		
	ACCESSION	AC093899	AC068884	
	VERSION	AC093899.3	GI:18497265	
	KEYWORDS	HTG.		
	SOURCE	human.		
	ORGANISM	Homo sapiens		
		Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;		
		Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.		

REFERENCE
AUTHORS
TITLE
JOURNAL
MEDLINE
PUBMED
REFERENCE
AUTHORS
TITLE
JOURNAL

1 (bases 1 to 172816)
Sultison,J.E. and Waterston,R.
Toward a complete human genome sequence
Genome Res. 8 (11), 1097-1108 (1998)
99063792
9847074
2 (bases 1 to 172816)
Pearman,C., Harkenson,W. and Boyer,E.
The sequence of Homo sapiens BAC clone RP11-724016
Unpublished (2001)
3 (bases 1 to 172816)
Waterston,R.H.
Direct Submission
Submitted (10-SEP-2001) Genome Sequencing Center, Washington
University School of Medicine, 4444 Forest Park Parkway, St. Louis,
MO 63108, USA
4 (bases 1 to 172816)
Waterston,R.H.
Direct Submission
Submitted (05-FEB-2002) Genome Sequencing Center, Washington
University School of Medicine, 4444 Forest Park Parkway, St. Louis,
MO 63108, USA
5 (bases 1 to 172816)
Waterston,R.
Direct Submission
Submitted (12-JUN-2002) Department of Genetics, Washington
University, 4444 Forest Park Avenue, St. Louis, Missouri 63108, USA
On Feb 5, 2002 this sequence version replaced g1:15625013.

COMMENT
JOURNAL
AUTHORS
TITLE
JOURNAL

Center: Washington University Genome Sequencing Center
Center code: WUGSC
Web site: http://genome.wustl.edu/gsc
Contact: saplens@wustl.wustl.edu
----- Summary Statistics -----
Drafting project name: H_NH0724016
Drafting Center: WMBR

NOTICE: This sequence may not represent the entire insert of this
clone. It may be shorter because we only sequence overlapping
clone sections once, or longer because we provide a small overlap
between neighboring data submissions.

This sequence was finished as follows unless otherwise noted:
all regions were double stranded, sequenced with an alternate
chemistry, or covered by high quality data (i.e., phred quality >=
30); an attempt was made to resolve all sequencing problems, such
as compressions and repeats; all regions were covered by sequence
from more than one subclone; and the assembly was confirmed by
restriction digest.

MAPPING INFORMATION:
Mapping information for this clone was provided by Dr. John D.
McPherson, Department of Genetics, Washington University, St. Louis
MO. For additional Information about the map position of this
sequence, see http://genome.wustl.edu/gsc

SOURCE INFORMATION:
The RC1C11 human BAC library was made from the blood of one male
donor, as described by Osogegawa,K., Moon,P.Y., Zhao,B., Frengen,E.,
Tateno,M., Catanesse,J.J. and de Jong,P.J. (1998) An improved
approach for construction of bacterial artificial chromosome
libraries. Genomics 51:1-8. The clone may be obtained either from
Research Genetics, Inc. (http://www.resgen.com) or Pieter de Jong
and coworkers at http://www.chori.org

VECTOR: pBACE3.6

NEIGHBORING SEQUENCE INFORMATION:
Actual start of this clone is at base position 1 of RP11-724016;
actual end is at base position 172816 of RP11-724016.

Unresolved simple sequence repeat from base position 22700 to
25900.

[illegible]

OY	1455	AATGATCTAAAGCATTGATTCCTTTGGGTACGTTCCGAAGCCCTCAAGCGCTGA	1514
Dd	24312	TATTAATAATATTTTATATTTATATACAGTTATATATAAATATTTATTTTATACAGTTAA	2437
OY	1515	ACCATCGCAACTATTTAAATTCGTTCTTTAATGCATCTTTAACATATTTATTGTAGTT	1574
Dd	24372	TAAATATATTTATTTTATTTTATACAGTTATATATA-ATATATTTAATTTTTTATACAGTTAAT	2443
OY	1575	GGAATTTAATAGACGGAACCTTGAACTTACATCAATTTTATATATACATCTAGTAGTA	1634
Dd	24431	AAATATATTTTATATTTTATACAGTTATATATAAATATTTATATTTTATACAGTTATATA	2449
OY	1635	TATTTCCAATATCATCTCTTGATGTTTAACTTAATCTTGTTCTTCTACGATATAA	1694
Dd	24491	ATA-TATTTATATTTTATATACAGTTATATATAAATATTTATGTTTATACATTTATATAA	2454
OY	1695	TATTTATATCCAGGTAAGAAAAAGTTTGCTTTATTTTTCGATGCAATGCAAGATATAAC	1754
Dd	24550	TATATTTATATTTTATATACATTTGTTTAAATATATATTT--ATATATAAATATATTTTA	2460
OY	1755	TATGACCTTTAATTTTGTGAAGAATGAATACCCCTTTACATCATGATTAATACCGTATGT	1814
Dd	24607	TATTTAATATATTTTATATATATAATATATATTTGATTTTAAATAATATTTATATATAA	2466
OY	1815	TTTGTTGCCATTAAGACAGCCTCTACAACCTGATGATGCAATTTTCTCGCAAAATTTAA	1874
Dd	24667	TATATGATTTTATTAATATGTTTAAATATAAATAATATATTTATATTAATATATATGTTTA	2472
OY	1875	ATTAGAAATTCATGCTACATATCAATAGAAAGAAAACAGCTGATTTACATTTTAATTTAA	1934
Dd	24727	TATATCATATATATATTTATATTTAATATATATTTTACATATATAATATATTTATATTTGA	2478
OY	1935	AGACAAATTTTTCGAAAAATGTTTAAATTTCTAACAAATATTTAAATATGATGCCAT	1994
Dd	24787	TATATATTTTATTTGATATATATATTTTATATATATATATATATTTTACATTTATATAT	2486
OY	1995	AATGATATTCCTATGTTCTTAAATATTTTATTTTATATTTAGTATTAATATACATTTGA	2054
Dd	24847	TTTATATATATTTTATATATATTTTATATATATATATATATATTTTATATATATTAATATAT	2496
OY	2055	ACCAAATA-AATGTGGGAATTCGAAATATCTCATTAATATTTTGGAAATCTACAATTT	2113
Dd	24907	TTTGATATATATTTTATATATATTTAATATTAATTAATTTATTTGATATATGTTTATATGATTAAT	2496
OY	2114	ATTATATTTTACTCAATACAAATGCATAGAAAAGTTCCAAAAAATTTTGTGTAACGAAA	2173
Dd	24967	ATATATTTTCGTTTATATATATATTTTATATGATTAATATATTTTACGTTTAAATATATTT	2506
OY	2174	CTTCCAAATTTTATTTTATATGGAACAGAAATAACAGATAGAAAACATTTTGTGTG	2233
Dd	25027	ATCTATTAATATATATTTTATGTTTATATATATTTTATATATTAATTAATATATTTACGTTTA	2508
OY	2234	GAATGGAGTACTAATATATACATTAAGCAAAATTTTAAAAAAT	2274
Dd	25087	ATATATATTTTATAGATAAATATATATTTTACGTTAATATAT	25127
RESULT 10			
PWMAL3P5			
LOCUS	PWMAL3P5	86827 bp	DNA
DEFINITION	Plasmodium falciparum MAL3P5, complete sequence.	linear	INV 04-MAY-2000
ACCESSION	AL034556 AL008971 AL008972 AL008978 AL010141 AL010153 AL010162		
VERSION	AL010206 AL010210 AL139179		
KEYWORDS	ALD034556.3 GI:7711064		
SOURCE	HTc; Centromere; CTRP protein; Initiation factor E4;		
ORGANISM	Setine/threonine protein phosphatase.		
REFERENCE	Plasmodium falciparum 3D7.		
AUTHORS	Bukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.		
	1 (bases 1 to 86827)		
	Churber, S., Lawson, D., Basham, D., Brown, D., Chillingworth, T.,		

[illegible]

Db	40328	TATATTATTAATATATATATTTTA-TTATTTATTAATTAATTTAATTTAATTTAATA	40386	TITLE AUTHORS JOURNAL REFERENCE	2 (bases 1 to 133877) Birren,B., Nusbaum,C. and Lander,E. Homo sapiens chromosome 18, clone CTD-2146H24
Db	40387	TTTAACTTTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA	40446	TITLE AUTHORS JOURNAL REFERENCE	2 (bases 1 to 133877) Birren,B., Linton,L., Nusbaum,C., Lander,E., All,A., Allen,N., Anderson,S., Barna,N., Baatlen,V., Bloom,T., Boguslavsky,L., Boukhalter,B., Brown,A., Camerata,J., Campolongo,A., Chang,J., Chazaro,B., Choepel,Y., Collanello,M., Collins,S., Collymore,A., Cook,A., Cooke,P., Dearellano,K., Dewar,K., Diaz,J.S., Dodge,S., Fero,S., Ferreira,P., Fitzhugh,W., Gage,D., Galagan,J., Gardyna,S., Ginde,S., Goid,S., Goyette,M., Graham,L., Grand-Pierre,N., Hagos,B., Horton,L., Hulme,W., Iliev,I., Johnson,R., Jones,C., Kamat,A., Karatas,A., Kells,C., Larocque,K., Lamares,R., Landers,T., Lehoczy,J., Levine,R., Lindblad-Toh,K., Liu,G., Maclean,C., Macdonald,P., Major,J., Marguis,N., Matthews,C., McCarthy,M., McEwan,P., McKernan,K., Meldrim,J., Menes,I., Mihova,T., Menga,V., Murphy,T., Naylor,J., Nguyen,C., Nicol,R., Nordu,C., Norman,C.H., O'Connor,T., O'Donnell,P., O'Neill,D., Oliver,J., Peterson,K., Phunkhang,P., Pterre,N., Pollara,V., Raymond,C., Retta,R., Riback,M., Riley,R., Rise,C., Rogov,P., Roman,J., Roselli,M., Roy,A., Santos,R., Schauer,S., Schupbeck,R., Seaman,S., Severy,P., Spencer,B., Stange-Thomann,N., Stojanovic,N., Strauss,N., Subramanian,A., Talamas,J., Testaye,S., Theodore,J., Topham,K., Travers,M., Travis,N., Triggillo,J., Vassiliev,H., Viel,R., Vo,A., Wilson,B., Wu,X., Wyman,D., Ye,W.D., Young,G., Zainoun,J., Zembek,L., Zimmer,A. and Zody,M.
Db	40627	TTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA	40686	TITLE AUTHORS JOURNAL REFERENCE	3 (bases 1 to 133877) Birren,B., Nusbaum,C., Lander,E., All,A., Allen,N., Anderson,S., Barna,N., Baatlen,V., Bloom,T., Boguslavsky,L., Boukhalter,B., Camerata,J., Chang,J., Chazaro,B., Choepel,Y., Collymore,A., Cook,A., Cooke,P., Dearellano,K., Dewar,K., Diaz,J.S., Dodge,S., Fero,S., Ferreira,P., Fitzhugh,W., Gage,D., Galagan,J., Gardyna,S., Goid,S., Ginde,S., Goyette,M., Graham,L., Grand-Pierre,N., Hagos,B., Horton,L., Hulme,W., Iliev,I., Johnson,R., Jones,C., Kamat,A., Karatas,A., Kells,C., Landers,T., Levine,R., Lindblad-Toh,K., Liu,G., Maclean,C., Macdonald,P., Major,J., Marguis,N., Matthews,C., McCarthy,M., McEwan,P., McKernan,K., Meldrim,J., Menes,I., Mihova,T., Menga,V., Murphy,T., Naylor,J., Nguyen,C., Nicol,R., Nordu,C., Norman,C.H., O'Connor,T., O'Donnell,P., O'Neill,D., Oliver,J., Peterson,K., Phunkhang,P., Pterre,N., Raymond,C., Retta,R., Rise,C., Rogov,P., Roman,J., Roselli,M., Roy,A., Santos,R., Schauer,S., Schupbeck,R., Seaman,S., Severy,P., Spencer,B., Stange-Thomann,N., Stojanovic,N., Strauss,N., Subramanian,A., Talamas,J., Testaye,S., Theodore,J., Topham,K., Travers,M., Travis,N., Triggillo,J., Vassiliev,H., Viel,R., Vo,A., Wilson,B., Wu,X., Wyman,D., Ye,W.D., Young,G., Zainoun,J., Zembek,L., Zimmer,A. and Zody,M.
Db	40657	ATTATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA	40686	TITLE AUTHORS JOURNAL REFERENCE	3 (bases 1 to 133877) Birren,B., Nusbaum,C., Lander,E., All,A., Allen,N., Anderson,S., Barna,N., Baatlen,V., Bloom,T., Boguslavsky,L., Boukhalter,B., Camerata,J., Chang,J., Chazaro,B., Choepel,Y., Collymore,A., Cook,A., Cooke,P., Dearellano,K., Dewar,K., Diaz,J.S., Dodge,S., Fero,S., Ferreira,P., Fitzhugh,W., Gage,D., Galagan,J., Gardyna,S., Goid,S., Ginde,S., Goyette,M., Graham,L., Grand-Pierre,N., Hagos,B., Horton,L., Hulme,W., Iliev,I., Johnson,R., Jones,C., Kamat,A., Karatas,A., Kells,C., Landers,T., Levine,R., Lindblad-Toh,K., Liu,G., Maclean,C., Macdonald,P., Major,J., Marguis,N., Matthews,C., McCarthy,M., McEwan,P., McKernan,K., Meldrim,J., Menes,I., Mihova,T., Menga,V., Murphy,T., Naylor,J., Nguyen,C., Nicol,R., Nordu,C., Norman,C.H., O'Connor,T., O'Donnell,P., O'Neill,D., Oliver,J., Peterson,K., Phunkhang,P., Pterre,N., Raymond,C., Retta,R., Rise,C., Rogov,P., Roman,J., Roselli,M., Roy,A., Santos,R., Schauer,S., Schupbeck,R., Seaman,S., Severy,P., Spencer,B., Stange-Thomann,N., Stojanovic,N., Strauss,N., Subramanian,A., Talamas,J., Testaye,S., Theodore,J., Topham,K., Travers,M., Travis,N., Triggillo,J., Vassiliev,H., Viel,R., Vo,A., Wilson,B., Wu,X., Wyman,D., Ye,W.D., Young,G., Zainoun,J., Zembek,L., Zimmer,A. and Zody,M.
Db	40667	ATTAAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA	40746	TITLE AUTHORS JOURNAL REFERENCE	3 (bases 1 to 133877) Birren,B., Nusbaum,C., Lander,E., All,A., Allen,N., Anderson,S., Barna,N., Baatlen,V., Bloom,T., Boguslavsky,L., Boukhalter,B., Camerata,J., Chang,J., Chazaro,B., Choepel,Y., Collymore,A., Cook,A., Cooke,P., Dearellano,K., Dewar,K., Diaz,J.S., Dodge,S., Fero,S., Ferreira,P., Fitzhugh,W., Gage,D., Galagan,J., Gardyna,S., Goid,S., Ginde,S., Goyette,M., Graham,L., Grand-Pierre,N., Hagos,B., Horton,L., Hulme,W., Iliev,I., Johnson,R., Jones,C., Kamat,A., Karatas,A., Kells,C., Landers,T., Levine,R., Lindblad-Toh,K., Liu,G., Maclean,C., Macdonald,P., Major,J., Marguis,N., Matthews,C., McCarthy,M., McEwan,P., McKernan,K., Meldrim,J., Menes,I., Mihova,T., Menga,V., Murphy,T., Naylor,J., Nguyen,C., Nicol,R., Nordu,C., Norman,C.H., O'Connor,T., O'Donnell,P., O'Neill,D., Oliver,J., Peterson,K., Phunkhang,P., Pterre,N., Raymond,C., Retta,R., Rise,C., Rogov,P., Roman,J., Roselli,M., Roy,A., Santos,R., Schauer,S., Schupbeck,R., Seaman,S., Severy,P., Spencer,B., Stange-Thomann,N., Stojanovic,N., Strauss,N., Subramanian,A., Talamas,J., Testaye,S., Theodore,J., Topham,K., Travers,M., Travis,N., Triggillo,J., Vassiliev,H., Viel,R., Vo,A., Wilson,B., Wu,X., Wyman,D., Ye,W.D., Young,G., Zainoun,J., Zembek,L., Zimmer,A. and Zody,M.
Db	40747	ATTATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA	40806	TITLE AUTHORS JOURNAL REFERENCE	3 (bases 1 to 133877) Birren,B., Nusbaum,C., Lander,E., All,A., Allen,N., Anderson,S., Barna,N., Baatlen,V., Bloom,T., Boguslavsky,L., Boukhalter,B., Camerata,J., Chang,J., Chazaro,B., Choepel,Y., Collymore,A., Cook,A., Cooke,P., Dearellano,K., Dewar,K., Diaz,J.S., Dodge,S., Fero,S., Ferreira,P., Fitzhugh,W., Gage,D., Galagan,J., Gardyna,S., Goid,S., Ginde,S., Goyette,M., Graham,L., Grand-Pierre,N., Hagos,B., Horton,L., Hulme,W., Iliev,I., Johnson,R., Jones,C., Kamat,A., Karatas,A., Kells,C., Landers,T., Levine,R., Lindblad-Toh,K., Liu,G., Maclean,C., Macdonald,P., Major,J., Marguis,N., Matthews,C., McCarthy,M., McEwan,P., McKernan,K., Meldrim,J., Menes,I., Mihova,T., Menga,V., Murphy,T., Naylor,J., Nguyen,C., Nicol,R., Nordu,C., Norman,C.H., O'Connor,T., O'Donnell,P., O'Neill,D., Oliver,J., Peterson,K., Phunkhang,P., Pterre,N., Raymond,C., Retta,R., Rise,C., Rogov,P., Roman,J., Roselli,M., Roy,A., Santos,R., Schauer,S., Schupbeck,R., Seaman,S., Severy,P., Spencer,B., Stange-Thomann,N., Stojanovic,N., Strauss,N., Subramanian,A., Talamas,J., Testaye,S., Theodore,J., Topham,K., Travers,M., Travis,N., Triggillo,J., Vassiliev,H., Viel,R., Vo,A., Wilson,B., Wu,X., Wyman,D., Ye,W.D., Young,G., Zainoun,J., Zembek,L., Zimmer,A. and Zody,M.
Db	40807	TATATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA	40864	TITLE AUTHORS JOURNAL REFERENCE	3 (bases 1 to 133877) Birren,B., Nusbaum,C., Lander,E., All,A., Allen,N., Anderson,S., Barna,N., Baatlen,V., Bloom,T., Boguslavsky,L., Boukhalter,B., Camerata,J., Chang,J., Chazaro,B., Choepel,Y., Collymore,A., Cook,A., Cooke,P., Dearellano,K., Dewar,K., Diaz,J.S., Dodge,S., Fero,S., Ferreira,P., Fitzhugh,W., Gage,D., Galagan,J., Gardyna,S., Goid,S., Ginde,S., Goyette,M., Graham,L., Grand-Pierre,N., Hagos,B., Horton,L., Hulme,W., Iliev,I., Johnson,R., Jones,C., Kamat,A., Karatas,A., Kells,C., Landers,T., Levine,R., Lindblad-Toh,K., Liu,G., Maclean,C., Macdonald,P., Major,J., Marguis,N., Matthews,C., McCarthy,M., McEwan,P., McKernan,K., Meldrim,J., Menes,I., Mihova,T., Menga,V., Murphy,T., Naylor,J., Nguyen,C., Nicol,R., Nordu,C., Norman,C.H., O'Connor,T., O'Donnell,P., O'Neill,D., Oliver,J., Peterson,K., Phunkhang,P., Pterre,N., Raymond,C., Retta,R., Rise,C., Rogov,P., Roman,J., Roselli,M., Roy,A., Santos,R., Schauer,S., Schupbeck,R., Seaman,S., Severy,P., Spencer,B., Stange-Thomann,N., Stojanovic,N., Strauss,N., Subramanian,A., Talamas,J., Testaye,S., Theodore,J., Topham,K., Travers,M., Travis,N., Triggillo,J., Vassiliev,H., Viel,R., Vo,A., Wilson,B., Wu,X., Wyman,D., Ye,W.D., Young,G., Zainoun,J., Zembek,L., Zimmer,A. and Zody,M.
Db	40865	TATATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA	40924	TITLE AUTHORS JOURNAL REFERENCE	3 (bases 1 to 133877) Birren,B., Nusbaum,C., Lander,E., All,A., Allen,N., Anderson,S., Barna,N., Baatlen,V., Bloom,T., Boguslavsky,L., Boukhalter,B., Camerata,J., Chang,J., Chazaro,B., Choepel,Y., Collymore,A., Cook,A., Cooke,P., Dearellano,K., Dewar,K., Diaz,J.S., Dodge,S., Fero,S., Ferreira,P., Fitzhugh,W., Gage,D., Galagan,J., Gardyna,S., Goid,S., Ginde,S., Goyette,M., Graham,L., Grand-Pierre,N., Hagos,B., Horton,L., Hulme,W., Iliev,I., Johnson,R., Jones,C., Kamat,A., Karatas,A., Kells,C., Landers,T., Levine,R., Lindblad-Toh,K., Liu,G., Maclean,C., Macdonald,P., Major,J., Marguis,N., Matthews,C., McCarthy,M., McEwan,P., McKernan,K., Meldrim,J., Menes,I., Mihova,T., Menga,V., Murphy,T., Naylor,J., Nguyen,C., Nicol,R., Nordu,C., Norman,C.H., O'Connor,T., O'Donnell,P., O'Neill,D., Oliver,J., Peterson,K., Phunkhang,P., Pterre,N., Raymond,C., Retta,R., Rise,C., Rogov,P., Roman,J., Roselli,M., Roy,A., Santos,R., Schauer,S., Schupbeck,R., Seaman,S., Severy,P., Spencer,B., Stange-Thomann,N., Stojanovic,N., Strauss,N., Subramanian,A., Talamas,J., Testaye,S., Theodore,J., Topham,K., Travers,M., Travis,N., Triggillo,J., Vassiliev,H., Viel,R., Vo,A., Wilson,B., Wu,X., Wyman,D., Ye,W.D., Young,G., Zainoun,J., Zembek,L., Zimmer,A. and Zody,M.
Db	40925	TTTAAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA	40984	TITLE AUTHORS JOURNAL REFERENCE	3 (bases 1 to 133877) Birren,B., Nusbaum,C., Lander,E., All,A., Allen,N., Anderson,S., Barna,N., Baatlen,V., Bloom,T., Boguslavsky,L., Boukhalter,B., Camerata,J., Chang,J., Chazaro,B., Choepel,Y., Collymore,A., Cook,A., Cooke,P., Dearellano,K., Dewar,K., Diaz,J.S., Dodge,S., Fero,S., Ferreira,P., Fitzhugh,W., Gage,D., Galagan,J., Gardyna,S., Goid,S., Ginde

[illegible]

On Jan 15, 2002 this sequence version replaced gi:18056753.

Summary Statistics -----

Sequencing vector: M13; 0%

Chemistry: Dye-Primer ET; 0% of reads

Chemistry: Dye-terminator Big Dye; 100% of reads

Assembly program: Phrap; version 0.990119

Consensus quality: 198878 bases at least Q40

Consensus quality: 199230 bases at least Q40

Consensus quality: 200198 bases at least Q30

Consensus quality: 200198 bases at least Q20

Insert size: 202000; agarose-fp

Insert size: 203204; sum-of-contigs

Quality coverage: 10.60 in Q20 bases; agarose-fp

Quality coverage: 10.87 in Q20 bases; sum-of-contigs

NOTE: this is a "working draft" sequence. It currently consists of 12 contigs. The true order of the pieces is not known and their order in this sequence record is arbitrary. Gaps between the contigs are represented as runs of N, but the exact sizes of the gaps are unknown. This record will be updated with the finished sequence as soon as it is available and the accession number will be preserved.

FEATURES	
*	1
*	1157
*	1257
*	2523
*	2524
*	2624
*	4637
*	4737
*	25188
*	23288
*	48385
*	48485
*	82313
*	82413
*	82414
*	122099
*	122199
*	200354
*	200454
*	201280
*	201380
*	202588
*	202689
*	203923
*	203924
*	204024
	1156: contig of 1156 bp in length
	1256: gap of unknown length
	2523: contig of 1267 bp in length
	2623: gap of unknown length
	4636: contig of 2013 bp in length
	4736: gap of unknown length
	25187: contig of 20451 bp in length
	23287: gap of unknown length
	48384: contig of 23097 bp in length
	48484: gap of unknown length
	82313: contig of 33829 bp in length
	82413: gap of unknown length
	122098: contig of 39685 bp in length
	122198: gap of unknown length
	200353: contig of 78155 bp in length
	200453: gap of unknown length
	201279: contig of 826 bp in length
	201379: gap of unknown length
	202588: contig of 1209 bp in length
	202688: gap of unknown length
	203923: contig of 1335 bp in length
	204023: gap of unknown length
	205130: contig of 1107 bp in length.

[illegible]

misc_feature	48485.	.82313	/note="assembly_name:Contig112"
misc_feature	82414.	.122098	/note="assembly_name:Contig113"
misc_feature	122199.	.200353	/note="assembly_name:Contig114"
misc_feature	200454.	.201279	/note="assembly_name:Contig17"
misc_feature	201380.	.202588	/note="assembly_name:Contig53"
misc_feature	202689.	.203923	/note="assembly_name:Contig81"
misc_feature	204024.	.205130	/note="assembly_name:Contig99"
misc_feature	41445	g 59044	t 1250 others
BASE COUNT	61131	a 42260	c 41445
ORIGIN			

Query Match	3.9%;	Score 124.6;	DB 2;	Length 205130;
Best Local Similarity	43.3%;	Pred. No. 2.5e-09;		
Matches 998;	Conservative	0;	Mismatches 1279;	Indels 26;

12/9; Anders 26; Gaps 8;

QY	128	ATTTAAATATTACAGTAATTAATTCCTCACTGATGTCTATAACAATAATCT	187
Db	79707	ATTATATATAAATATATAAATATGTAATATATTTTATATATATATATATTAATAATAT	79648

QY 188 A C T T T G G A T A T G A C A A A A T A T G A G T T T G C A T T T A T A A T A A A G A A T A A T C G A 247
 || ||||| | | ||||| ||| | | |||
 Db 79647 T A T A T T A T A T A T A T T A T A T A T A T A T A T C A T A T T A T A T T C A C A T A T A T A T --- 79591

QY 248 TTCACATTTGGTTGGATTACACAGTTAAGTTTGTGTCTTCTTTTGTATATGATATGAG 307
 || || | | |||| | | || | || |||| ||||
 Db 79590 TGTATATTATATAAATACATATTATATATTATATATTATATATATAAATCAAA 79531

QY 308 TAAATCAAAAAGATGATTGATTGAAGTGAACATATTTCTGTATGACCCCAAAAAA 367
|| || | | | | | | | | | | | | | |
Db 79530 TATATATATTTATATATATATAACTTATATTTATATAACAATTAATCATATATAT 79471

QY 368 AAAAAAAAAAACAACAAACCCGCCGATATAGTCTTTGGTTCGATTAGCTTTA 427
| | | | |
79470 TATATTATATGACATATAAATACATATTATATATATATTAATTAACATTAATACATA 79411
| | | | |

QY 428 TTGTGATCAATAATTACATGCATCTTCCTTGATTAAGTAAAGATTTCTTACCATAA 487
|| || | | | | | | | | | | | | | | | |
Db 79410 TTAATATTATATTAATTACATATAAATACATATATTAATTCATTATATACAAAATAC 79351

488 AATTTCGAAATCATATCTCTGATTTAAATTAATACGAGTGGAAATATCCGTTATC 547
 79350 ATATATATTATATACATAAATACATATATATTTATATACCTAAATACATATTTGATATTA 79291

548 GATCAGCTCCAAATCATGATTTATGATTTCTGTGCTAACTCCAGAAATTATTAACAAGAGTAT 607
 79290 TTTCATATAATATATACATATTGTAATATATTTTATATAGATATA---TCAATATTTGAT 79335

608 TGAGAAAAACCGAAATAGAAAAAGGAAAGACTAGTACCATGAGTATGCAATTA 667
 79234 ATTGCATTTCATTTATATATACATATTTGTAATTTATATATATATATATAATCAATTCATTTAT 79175

668 TTATCAAGGACATTAAGGATGACACCAAAAGGTGTGCAATATGTCCTCCAGCT 727
||||| - - - - -
79174 TTATATTTATATAATACATATCATATATTTAATATATATAACATATCATATTAATTT 79115

728 TTCCTGACATCATATCGACCTATTGGATTTTCGATATCTGTTAAATTTGCGA 787
79114 ATATTATTANGTAAATATATATGACATATATTATTTTATGTTATATATATTTCA 79055

788 TAACGATGTGAAAAAATTATAATTGGTAGCTGCATCTCAATTAATGTCACAGTA--- 844
|| ||| | || ||| | || ||| ||
79054 TATATATTAATTATATATATATATATATATATCATATATATTTTAAATTTTATATACG 79095

[illegible]

Db 3941 ATTATGATATTTATATATATATATAT -ATATAGAAAAATTAATTTATTAATTA 3999
 Qy 1787 TTTACTCATATATATATACCGATGTTTGGTCCATATAGACGCTCACAACGT 1846
 Db 4000 TTTAT 4059
 Qy 1847 GATAGTCAATTTTTCGCAATATATATATATATATATATATATATATATATAT 1906
 Db 4060 AATTAATCATTTTATATATATATATATATATATATATATATATATATATAT 4119
 Qy 1907 AACAGCTGAT 1966
 Db 4120 AAT 4179
 Qy 1967 AACAT 2026
 Db 4180 TTTTAAAAAAACCTATATATATATATATATATATATATATATATATATAT 4239
 Qy 2027 TTTTAT 2086
 Db 4240 TTTAAT 4299
 Qy 2087 ATTAAT 2146
 Db 4300 AT 4359
 Qy 2147 TTTCAAAAAATTTTGTGAAGAACTCCAAATTTTATATATATATATATATAT 2206
 Db 4360 TAAAT 4419
 Qy 2207 TTAACAGTAT 2266
 Db 4420 TAT 4479
 Qy 2267 TAAATAAT 2326
 Db 4480 AAAAAATTTCTTAATATATATATATATATATATATATATATATATATATAT 4539
 Qy 2327 ATGCATGTCGATTCAGAAATGGACAAATGAAACGAAATTAATTAATCTTA 2386
 Db 4540 TTTTAAAAAAATAGTAAATATATATATATATATATATATATATATATAT 4599

RESULT 15
 DMU37541 19517 bp DNA circular INV 02-MAR-2001
 LOCUS Drosophila melanogaster complete mitochondrial genome.
 DEFINITION U37541
 ACCESSION U37541
 VERSION U37541.1 GI:1166529
 KEYWORDS
 SOURCE Drosophila melanogaster.
 ORGANISM Drosophila melanogaster
 Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
 Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
 Ephydroidea; Drosophilidae; Drosophila.
 1 (bases 12511 to 12682)
 Clary,D.O., Goddard,J.M., Martin,S.C., Faaron,C.M. and
 Wolstenholme,D.R.
 Drosophila mitochondrial DNA: a novel gene order
 Nucleic Acids Res. 10 (21), 6619-6637 (1982)
 2 (bases 5269 to 5695)
 Clary,D.O., Wahlsteiner,J.A. and Wolstenholme,D.R.
 Transfer RNA genes in Drosophila mitochondrial DNA: related 5'
 flanking sequences and comparisons to mammalian mitochondrial tRNA
 genes
 Nucleic Acids Res. 11 (8), 2411-2425 (1983)
 JOURNAL 83220794
 MEDLINE 6304652
 PUBMED 3 (bases 404 to 5272)
 REFERENCE de Bruijn,M.H.
 TITL Drosophila melanogaster mitochondrial DNA, a novel organization and

JOURNAL genetic code
 MEDLINE Nature 304 (5923), 234-241 (1983)
 PUBMED 83245048
 4 (bases 804 to 1778)
 REFERENCE Satta,Y., Ishiwa,H. and Chigusa,S.I.
 ANALYSIS OF NUCLEOTIDE SUBSTITUTIONS OF MITOCHONDRIAL DNAs IN
 Drosophila melanogaster and its sibling species
 Mol. Biol. Evol. 4 (6), 638-650 (1987)
 JOURNAL 88174373
 MEDLINE 2832697
 PUBMED 5 (bases 5268 to 13619)
 REFERENCE Garesse,R.
 Drosophila melanogaster mitochondrial DNA: gene organization and
 evolutionary considerations
 Genetics 118 (4), 649-663 (1988)
 JOURNAL 88212147
 MEDLINE 3130291
 PUBMED 6 (bases 441 to 2967)
 REFERENCE Satta,Y. and Takahata,N.
 EVOLUTION OF DROSOPHILA MITOCHONDRIAL DNA AND THE HISTORY OF THE
 MELANOGASTER SUBGROUP
 Proc. Natl. Acad. Sci. U.S.A. 87 (24), 9558-9562 (1990)
 JOURNAL 91088557
 MEDLINE 2124697
 PUBMED 7 (bases 14215 to 14512)
 REFERENCE Ballard,J.W., Olsen,G.J., Faith,D.P., Odgers,W.A., Rowell,D.M. and
 Atkinson,P.W.
 EVIDENCE FROM 12S RIBOSOMAL RNA SEQUENCES THAT ONYCHOPHORANS ARE
 MODIFIED ARTHROPODS
 Science 258 (5086), 1345-1348 (1992)
 JOURNAL 1455227
 MEDLINE 8 (bases 14917 to 19517)
 PUBMED 93088057
 REFERENCE Lewis,D.L., Farr,C.L., Farquhar,A.L. and Kaguni,L.S.
 SEQUENCE, ORGANIZATION, AND EVOLUTION OF THE A+T REGION OF
 DROSOPHILA MELANOGASTER MITOCHONDRIAL DNA
 Mol. Biol. Evol. 11 (3), 523-538 (1994)
 JOURNAL 94285822
 MEDLINE 8015445
 PUBMED 9 (bases 1 to 408; 13319 to 19517)
 REFERENCE Lewis,D.L., Farr,C.L. and Kaguni,L.S.
 DROSOPHILA MELANOGASTER MITOCHONDRIAL DNA: COMPLETION OF THE
 NUCLEOTIDE SEQUENCE AND EVOLUTIONARY COMPARISONS
 Insect Mol. Biol. 4 (4), 263-278 (1995)
 JOURNAL 96423163
 MEDLINE 8825764
 PUBMED 10 (bases 1 to 19517)
 REFERENCE Lewis,D.L., Farr,C.L. and Kaguni,L.S.
 DIRECT SUBMISSION
 Submitted (03-OCT-1995) Laurie S. Kaguni, Biochemistry Department,
 State University, East Lansing, MI 48824-1319, USA
 Michigan location/Qualifiers
 1. 19517
 /organism="Drosophila melanogaster"
 /organelle="mitochondrion"
 /db_xref="taxon:722"
 /note="derived from new and previously submitted
 sequences; sequence is a composite containing sequences
 obtained from different Drosophila melanogaster strains"
 1. 65
 /product="tRNA-Tle"
 /complement(97. 165)
 /product="tRNA-Gln"
 171. 239
 /product="tRNA-Phe"
 240. 1265
 /codon_start=1
 /transl_table=5
 /product="NADH dehydrogenase subunit 2"
 /protein_id="AAC47811.1"
 /db_xref="GI:1166530"
 /translation="MFNNSKILFTITMIGTITVTSNSWLGAMGLEINLSFIPL

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OW nucleic - nucleic search, using sw model

Run on: June 10, 2003, 21:59:33 ; Search time 733.005 Seconds
(without alignments)
9837.455 Million cell updates/sec

Title: US-09-502-426B-1_COPY_1_3202
Perfect score: 3202
Sequence: 1 atgtggtattatattgtt.....gagagagagaactagctcc 3202

Scoring table: IDENTITY NUC
Gapop 10.0, Gapext 1.0

Searched: 2185239 seqs, 112599159 residues

Total number of hits satisfying chosen parameters: 4370478

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database :
N.GeneSeq_101002:*

- 1: /SIDS2/gcgdata/geneSeq/geneSeq-emb1/NA1980.DAT:*
- 2: /SIDS2/gcgdata/geneSeq/geneSeq-emb1/NA1981.DAT:*
- 3: /SIDS2/gcgdata/geneSeq/geneSeq-emb1/NA1982.DAT:*
- 4: /SIDS2/gcgdata/geneSeq/geneSeq-emb1/NA1983.DAT:*
- 5: /SIDS2/gcgdata/geneSeq/geneSeq-emb1/NA1984.DAT:*
- 6: /SIDS2/gcgdata/geneSeq/geneSeq-emb1/NA1985.DAT:*
- 7: /SIDS2/gcgdata/geneSeq/geneSeq-emb1/NA1986.DAT:*
- 8: /SIDS2/gcgdata/geneSeq/geneSeq-emb1/NA1987.DAT:*
- 9: /SIDS2/gcgdata/geneSeq/geneSeq-emb1/NA1988.DAT:*
- 10: /SIDS2/gcgdata/geneSeq/geneSeq-emb1/NA1989.DAT:*
- 11: /SIDS2/gcgdata/geneSeq/geneSeq-emb1/NA1990.DAT:*
- 12: /SIDS2/gcgdata/geneSeq/geneSeq-emb1/NA1991.DAT:*
- 13: /SIDS2/gcgdata/geneSeq/geneSeq-emb1/NA1992.DAT:*
- 14: /SIDS2/gcgdata/geneSeq/geneSeq-emb1/NA1993.DAT:*
- 15: /SIDS2/gcgdata/geneSeq/geneSeq-emb1/NA1994.DAT:*
- 16: /SIDS2/gcgdata/geneSeq/geneSeq-emb1/NA1995.DAT:*
- 17: /SIDS2/gcgdata/geneSeq/geneSeq-emb1/NA1996.DAT:*
- 18: /SIDS2/gcgdata/geneSeq/geneSeq-emb1/NA1997.DAT:*
- 19: /SIDS2/gcgdata/geneSeq/geneSeq-emb1/NA1998.DAT:*
- 20: /SIDS2/gcgdata/geneSeq/geneSeq-emb1/NA1999.DAT:*
- 21: /SIDS2/gcgdata/geneSeq/geneSeq-emb1/NA2000.DAT:*
- 22: /SIDS2/gcgdata/geneSeq/geneSeq-emb1/NA2001A.DAT:*
- 23: /SIDS2/gcgdata/geneSeq/geneSeq-emb1/NA2001B.DAT:*
- 24: /SIDS2/gcgdata/geneSeq/geneSeq-emb1/NA2002.DAT:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	3202	100.0	6888	21	AAA59599
2	84.6	2.6	883	22	AAA15210
3	84.6	2.6	9810	24	ABK32427
4	83.8	2.6	8170	24	ABK28257
5	82.8	2.6	6131	24	ABL32890
6	82.6	2.6	9810	24	ABL32426
7	82.6	2.6	15732	22	AAA54388
8	82.6	2.6	15732	24	ABK28233
9	81.6	2.5	5689	22	AAA54384

C	10	81.6	2.5	5689	22	AAA546426	Tumour suppressor
C	11	81.6	2.5	5689	24	ABK28226	DNA transcription
C	12	80.6	2.5	6294	24	ABL33054	Human immune syste
C	13	80.4	2.5	6131	24	ABL32891	Human immune syste
C	14	80	2.5	64081	24	ABA32787	Buchnera sp. genom
C	15	79.6	2.5	47108	24	ABK31511	Signal transductio
C	16	79.4	2.5	883	22	AAA15210	Human breast cance
C	17	79.4	2.5	8310	20	AAZ29911	CDNA encoding a SC
C	18	79.2	2.5	18283	24	ABL70502	Chemically treated
C	19	79.2	2.5	18283	24	AA561363	Human gene regulat
C	20	78.2	2.4	50000	24	ABL5643	AMEPV genome fragm
C	21	77.8	2.4	17137	24	ABL32191	Human immune syste
C	22	77.6	2.4	9810	24	ABL32427	Human immune syste
C	23	77.4	2.4	19124	18	AA728882	Plasmodium var-7 g
C	24	77.4	2.4	19124	21	AA298287	Plasmodium var-7 p
C	25	77.2	2.4	15373	24	ABL32467	Human immune syste
C	26	77	2.4	11691	24	ABL34241	Human immune syste
C	27	76.8	2.4	50000	24	ABL5643	AMEPV genome fragm
C	28	76.4	2.4	9810	24	ABL32426	Human immune syste
C	29	76.2	2.4	10254	24	ABL33075	Human immune syste
C	30	76.2	2.4	15548	24	ABL34155	Human immune syste
C	31	76	2.4	50000	24	ABL56202	AMEPV genome fragm
C	32	75.6	2.4	6050	24	ABL34011	Human immune syste
C	33	75.2	2.3	5979	22	AAA54513	Human immune syste
C	34	75.2	2.3	5979	24	ABK28152	Chemically pretrea
C	35	75	2.3	38342	22	AAA6746	DNA transcription
C	36	75	2.3	38342	24	ABK31507	Tumour suppressor
C	37	74.8	2.3	6963	24	ABL32979	Signal transductio
C	38	74.8	2.3	10710	24	ABL3893	Human immune syste
C	39	74.8	2.3	11422	24	ABK39936	Human immune syste
C	40	74.8	2.3	11422	24	ABL33218	Human immune syste
C	41	74.4	2.3	10250	24	ABN80083	Human chemically m
C	42	74.2	2.3	700	22	AAH93026	Human inflammatory
C	43	74.2	2.3	6641	24	ABL54336	Chemically treated
C	44	74.2	2.3	6641	24	ABL33315	Human immune syste
C	45	74	2.3	2104	13	AAQ25273	Sequence encoding

ALIGNMENTS

RESULT 1
AAA59599
ID AAA59599 standard; DNA; 6888 BP.

14-NOV-2000 (first entry)

DNA encoding a cytochrome P450 enzyme designated DMF4.

DMF4; cytochrome P450 enzyme; brassinosteroid; 22alpha-hydroxylation;
plant phenotype; cell elongation; ss.

Arabidopsis sp.

OS	XX	Key	Location/Qualifiers
OS	XX	promoter	1..3203
FT	FT	TATA_signal	3060..3125
FT	FT	CDS	3203..6110
FT	FT	exon	3203..3423
FT	FT	intron	3424..3503
FT	FT	exon	3504..3828
FT	FT	intron	3829..3913

FT exon 3914..4066
 FT /tag- h
 FT Intron 4067..4164
 FT /tag- i
 FT exon 4165..4479
 FT /tag- j
 FT Intron 4480..4631
 FT /tag- k
 FT exon 4632..4724
 FT /tag- l
 FT Intron 4725..4815
 FT /tag- m
 FT exon 4816..4894
 FT /tag- n
 FT Intron 4895..5000
 FT /tag- o
 FT exon 5001..5110
 FT /tag- p
 FT Intron 5111..5864
 FT /tag- q
 FT exon 5865..6110
 FT /tag- r
 FT 3' UTR 6011..6468
 FT /tag- s
 PN W0200047715-A2.
 PD 17-AUG-2000.
 PE 11-FEB-2000; 2000WO-US03820.
 PR 11-FEB-1999; 99US-0119657.
 PR 11-FEB-1999; 99US-0119658.
 XX (ARIZ-) ARIZONA BOARD OF REGENTS.
 PA Aspiroz R, Choe S, Feldmann KA;
 PI WPI: 2000-549142/50.
 DR P-PSDB; AAB07921.
 XX
 PT New isolated dwf4 polynucleotide useful for altering the phenotype of
 PT plants, for diagnostic assays and in the production of antibodies -
 PS Claim 3; Fig 10A-G; 113pp; English.
 XX
 CC The present sequence encodes a DMF4 polypeptide. The polypeptide is a
 CC cytochrome P450 enzyme that mediates multiple steps in synthesis of
 CC brassinosteroids. Specifically, it mediates multiple
 CC 22alpha-hydroxylation steps in brassinosteroid biosynthesis. The DMF4
 CC polynucleotide is used for altering the phenotype of a plant. DMF4
 CC plants display a dramatic reduction in the length of different organs,
 CC and this size reduction is attributable to a defect in cell elongation.
 CC The DMF4 polynucleotides and polypeptides can be used in diagnostic
 CC assays and to generate antibodies, which can be used to produce
 CC immunogenic compositions.
 XX
 SQ Sequence 6888 BP; 2294 A; 1010 C; 1193 G; 2391 T; 0 other;
 Query Match 100.0%; Score 3202; DB 21; Length 6888;
 Best Local Similarity 100.0%; Pred. No. 0;
 Matches 3202; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Db 121 AAAATTCATTAAAAATTTTACAGTAATTAATTAATCTTACATGTATTTATTAACAA 180
 Qy 181 AATATCTAATCTTGGTATATGAGAAATATGAGTTTGAATTTAATATAAGGANA 240
 Db 181 AATATCTAATCTTGGTATATGAGAAATATGAGTTTGAATTTAATATAAGGANA 240
 Qy 241 TAATGATTCATTTGGTTGATTAACAGATTAAGTTTGTGTCTTTGTATATGT 300
 Db 241 TAATGATTCATTTGGTTGATTAACAGATTAAGTTTGTGTCTTTGTATATGT 300
 Qy 301 ATATGATTAATCAAAAAAGATATGATTTGAAGTGAACATATTCGTATAGCCCCCA 360
 Db 301 ATATGATTAATCAAAAAAGATATGATTTGAAGTGAACATATTCGTATAGCCCCCA 360
 Qy 361 AAAAAAAAAAAAAACAACAACACACCCCCCGATATAGTTTGTGTGTGAT 420
 Db 361 AAAAAAAAAAAAAACAACAACACACCCCCCGATATAGTTTGTGTGTGAT 420
 Qy 421 AGTTTATTTGATCAATTAATACATCATTTCTTGTATATCAATGAGATTTCTTAC 480
 Db 421 AGTTTATTTGATCAATTAATACATCATTTCTTGTATATCAATGAGATTTCTTAC 480
 Qy 481 CAATTAATTTGCAATTCATATCTCTGATTTATTAATTAATACAGATGATATTC 540
 Db 481 CAATTAATTTGCAATTCATATCTCTGATTTATTAATTAATACAGATGATATTC 540
 Qy 541 GTTATGATCACTCCATCATGATTAATGATTTCTGTGCTAATCCAGCAATTAATACA 600
 Db 541 GTTATGATCACTCCATCATGATTAATGATTTCTGTGCTAATCCAGCAATTAATACA 600
 Qy 601 AGAGTATGAGAAAAACCAAAATTAAGAAAAAGAGATGATGACCATGAGATATG 660
 Db 601 AGAGTATGAGAAAAACCAAAATTAAGAAAAAGAGATGATGACCATGAGATATG 660
 Qy 661 TGAATTAATTAACAAAGAAATTAAGATGACACCAAAAGTGTGGAATATGTCCT 720
 Db 661 TGAATTAATTAACAAAGAAATTAAGATGACACCAAAAGTGTGGAATATGTCCT 720
 Qy 721 GCCAGCTTCTCTCAATCAATATGACACCCATTTGATTTCTGATATTCGTTAAA 780
 Db 721 GCCAGCTTCTCTCAATCAATATGACACCCATTTGATTTCTGATATTCGTTAAA 780
 Qy 781 TTGGGATTAACGATTTGTAAGAAATATTTTATTTTGTACCTGATATATATGTTCA 840
 Db 781 TTGGGATTAACGATTTGTAAGAAATATTTTATTTTGTACCTGATATATATGTTCA 840
 Qy 841 GGTATTTTCATTAATCTTCTGTTTAAAGATATTTGCTTTTGTGCTTCTCT 900
 Db 841 GGTATTTTCATTAATCTTCTGTTTAAAGATATTTGCTTTTGTGCTTCTCT 900
 Qy 901 TAACATATTAATTAATCGGATATATGATTAACATGATATATCAACAAATATGCTGG 960
 Db 901 TAACATATTAATTAATCGGATATATGATTAACATGATATATCAACAAATATGCTGG 960
 Qy 961 ACCATTTGGAATTAACCTTTTCTCAACATTAACGGAACATGACCTTCAATTA 1020
 Db 961 ACCATTTGGAATTAACCTTTTCTCAACATTAACGGAACATGACCTTCAATTA 1020
 Qy 1021 CGATTTTACAGCGTACATGATTAAGATTAAGATTAAGCAACCCGCTCAAGC 1080
 Db 1021 CGATTTTACAGCGTACATGATTAAGATTAAGATTAAGCAACCCGCTCAAGC 1080
 Qy 1081 TATTTATTAACAAGTTTACAACTGAATATAGCTTGAATATTCCTTGAATTA 1140
 Db 1081 TATTTATTAACAAGTTTACAACTGAATATAGCTTGAATATTCCTTGAATTA 1140
 Qy 1141 CCGGTTGTTATGTAATATATAGATTTAGTGAACAAATATGTTAATCAATAGTGCTCA 1200
 Db 1141 CCGGTTGTTATGTAATATATAGATTTAGTGAACAAATATGTTAATCAATAGTGCTCA 1200
 Qy 1201 ACATATACATTAATCTTACAGAAAAAACAACCTTAAGAGAGTTAACATATCATATAT 1260
 Db 1201 ACATATACATTAATCTTACAGAAAAAACAACCTTAAGAGAGTTAACATATCATATAT 1260

QY	2341	TCGAGATTGGGACAACAATGAAAAACGAATTAAATATTACTTTAAATTAATTAATTAAT	2400
Db	2341	TCGAGAAATGGGACAACAATGAAAAACGAATTAAATTAATTAATTAATTAATTAAT	2400
QY	2401	TTGAGTAAATGTTGTTTCTGACTATTTGAGGGGCAAAAAAAGACATGCCAAAAGTCTAC	2460
Db	2401	TTGATTAATATGTTTCTGACTATTTGAGGGGCAAAAAAAGACATGCCAAAAGTCTAC	2460
QY	2461	GGGTTGACTGTCCAGTTCCGTAATATCTAATTAATCTGCTCTTTGACCGCAGCTCGTG	2520
Db	2461	GGGTTGACTGTCCAGTTCCGTAATATCTAATTAATCTGCTCTTTGACCGCAGCTCGTG	2520
QY	2521	TAGGGGTCCTTCACATTTTCTACTGTTTACCCCTACTGTGAGCCACCCCTTTCCCA	2580
Db	2521	TAGGGGTCCTTCACATTTTCTACTGTTTACCCCTACTGTGAGCCACCCCTTTCCCA	2580
QY	2581	TATCTAAGGGTAATTTTGGAAATGCCAATTTAACCAGATGAGACCGTACCGGACTTCC	2640
Db	2581	TATCCCTAAGGGTAATTTTGGAAATGCCAATTTAACCAGATGAGACCGGACTTCC	2640
QY	2641	TGGGATTCCTCGGAGCATTTATCAAAAATTTATAGACAGAAATGGGTTATTAATTTAA	2700
Db	2641	TGGGATTCCTCGGAGCATTTATCAAAAATTTATAGACAGAAATGGGTTATTAATTTAA	2700
QY	2701	AACTCAAACTTGATCAGATTAATAATTTCTAATAACCTTTACGATGATGTCAGCATCT	2760
Db	2701	AACTCAAACTTGATCAGATTAATAATTTCTAATAACCTTTACGATGATGTCAGCATCT	2760
QY	2761	ANCTATGACTTTTTTTTTTCTACACAGGTGAGTAAAGTTATAGTACTATTAGCCAGAG	2820
Db	2761	ANCTATGACTTTTTTTTTTCTACACAGGTGAGTAAAGTTATAGTACTATTAGCCAGAG	2820
QY	2821	ACAATTGATATAGATATATCCATTAATCCATGATTTATGATATTAATAGCTGTAAA	2880
Db	2821	ACAATTGATATAGATATATCCATTAATCCATGATTTATGATATTAATAGCTGTAAA	2880
QY	2881	CTATTTGAGCATGCGAGCTTCTGCCACTTTTGTTTTAAATTTAAGATTTAATAATA	2940
Db	2881	CTATTTGAGCATGCGAGCTTCTGCCACTTTTGTTTTAAATTTAAGATTTAATAATA	2940
QY	2941	AAGTATTAAGAAGGCAATTAACGAGGCAACAAAGTAAATGACACGGAGAAACAAAAGCCA	3000
Db	2941	AAGTATTAAGAAGGCAATTAACGAGGCAACAAAGTAAATGACACGGAGAAACAAAAGCCA	3000
QY	3001	TGAAGCTCATGTTGTTTAACTTAATTAAGAAATTTTAAATTTTAAATGAGCATG	3060
Db	3001	TGAAGCTCATGTTGTTTAACTTAATTAAGAAATTTTAAATTTTAAATGAGCATG	3060
QY	3061	ATAACAATATATATTTCTGACTTCTTAAACCCCTCTTACAAACAGAAAGTCCCTTTT	3120
Db	3061	ATAACAATATATATTTCTGACTTCTTAAACCCCTCTTACAAACAGAAAGTCCCTTTT	3120
QY	3121	TCAGTAGAAGTCGATTCCTCAATCTTAAAGACAAAAGCCATTGAAAGAAAGTAGATGA	3180
Db	3121	TCAGTAGAAGTCGATTCCTCAATCTTAAAGACAAAAGCCATTGAAAGAAAGTAGATGA	3180
QY	3181	GAGAGAGACAGAAACTAGCTCC	3240
Db	3181	GAGAGAGACAGAAACTAGCTCC	3240
RESULT 2	AA15210/c		
xx	AA15210	standard; cDNA: 883 BP.	
xx	AA15210;		
AC	07-DEC-2001	(first entry)	
xx			
DE	Human breast cancer expressed polynucleotide 7667.		
xx			
xx	Human; breast cancer; cell marker; cytosolic; ss.		

OS Homo sapiens.
 XX MO200151628-A2.
 XX 19-JUL-2001.
 PD 10-JAN-2001; 2001WO-US00798.
 XX 14-JAN-2000; 2000US-0176077.
 PR 14-MAR-2000; 2000US-0189167.
 PR 24-MAR-2000; 2000US-0192089.
 PR 29-MAR-2000; 2000US-0193480.
 PR 15-MAY-2000; 2000US-0205230.
 PR 09-JUN-2000; 2000US-0211315.
 PR 25-JUL-2000; 2000US-0220534.
 XX (MILL-) MILLENNIUM PREDICTIVE MEDICINE INC.
 PA L1111e J, Xu Y, Wang Y, Steinmann K;
 PI WPI; 2001-451856/48.
 DR New peptide useful as a marker for the diagnosis of breast cancer -
 PT Claim 1; Page 1378; 3695pp; English.
 XX The invention relates to human breast cancer expressed polynucleotides
 CC (AL07544-AL26789) and methods of assessing whether a patient is
 CC afflicted with breast cancer by examining the correlation between the
 CC expression of certain markers and the cancerous state of breast cells.
 CC The polynucleotides and encoded polypeptides are potential markers for
 CC detecting, diagnosing, monitoring, characterizing, treating and
 CC potentially preventing breast cancer. The polynucleotides and encoded
 CC polypeptides are also useful for isolating compounds with cytostatic
 CC activity.
 XX
 XX Sequence 883 BP; 322 A; 18 C; 23 G; 396 T; 124 other;
 SO
 Query Match 2.6%; Score 84.6; DB 22; Length 883;
 Best Local Similarity 39.1%; Pred. No. 0.00015;
 Matches 347; Conservative 0; Mismatches 533; Indels 7; Gaps 1;

OY 1909 CAGCTGAGTATTACATTTTAAATTTAAAGCAAAATTTTGAAGAAATGTTATATTCTTA 1968
 DB 410 AANANNAATTAATTTTNNAAAAATTTTATNTATNTANNNANANNAATTTTAAAN 351
 OY 1969 CAATATTATTAATGATGCTATATATGTTTCCATGTTCTTAATATTTT 2028
 DB 350 TNAANANNAATTTTANTTTTAAATTAATTAATTAATTAATTAATTTTAA 291
 OY 2029 TATATTGTTTAAATACATTTATGACCAATTAATAGTTGTAATTCATTCAT 2088
 DB 290 TATTTAAAAAAATTTTATTTATATTAATTTTATTTTAAAAATTTTAAAN 231
 OY 2089 TAAATTTTGAATTCATCAATATTAATTTAGTCAATTAACATTCATAGAAAGT 2148
 DB 230 TAAATTTATTAAGNNNNNTNATTTTAAAAATTAATTTNNAATTTTAAATTTG 171
 OY 2149 CCAAAAAAATTTTGTACAGAACTTCCAAATTTTTTTATTTATGACACAGAAATA 2208
 DB 170 NAAAAAATTTATTAATTAATTTTATTTTAAACNATGTAATTAATTAANANA 111
 OY 2209 ACAGATGAAAACTATTTTGTGGAATGGAATGATTAATTAATTAATTAATTA 2268
 DB 110 ATAGNATTTTGAATTTAAATTTTAAATATCTTTTNAATTTAAATTTTAA 51
 OY 2269 AAAAAATTAATAGCTATACGCGCTCAAGATATGTTATCTAGTAGG 2315
 DB 50 AAAAAAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 4
 RESULT 3
 ABL32427
 ID ABL32427 standard; DNA; 9810 BP.
 AC ABL32427;
 XX 26-MAR-2002 (first entry)
 DE Human immune system associated gene SEQ ID NO: 400.
 XX Human: immune system disease; cytosine methylation; antiasthmatic;
 KW antidiabetic; antineoplastic; antineoplastic; antineoplastic;
 KW antineoplastic; antineoplastic; antineoplastic; antineoplastic;
 KW antineoplastic; antineoplastic; antineoplastic; antineoplastic;
 KW acute myeloid leukaemia; Alzheimer's disease; AIDS; epilepsy;
 KW neurofibromatosis; rheumatoid arthritis; psoriasis; bowel disease;
 KW gene; ds.
 XX Homo sapiens.
 OS WO200200928-A2.
 PN 03-JAN-2002.
 PD 02-JUL-2001; 2001WO-EP07537.
 XX 30-JUN-2000; 2000DE-1032529.
 PR 01-SEP-2000; 2000DE-1043826.
 XX (EPIG-) EPIGENOMICS AG.
 PA Olek A, Piepenbrock C, Berlin K;
 PI WPI; 2002-130909/17.
 DR Nucleic acid comprising fragment of chemically modified gene, useful
 PT for diagnosis and treatment of diseases associated with abnormal
 PT cytosine methylation -
 XX Claim 1; SEQ ID NO 400; 32pp + Sequence Listing; German.
 XX The present invention provides a number of human immune system associated

QY 1937 ACAAATTTTGGAAAAAGCTTATATCTTCAACAATATTAATAATGATGCGCTATTA 1996
 Db 488 TATTTTGTATTAATTAATGATGCTTATATTTTGTGTAATTTTGTGAAAT 547
 QY 1997 TGTATTTCCATGTTCTTAAATATTTTATTTTATTTTATTTTATTTTAAAC 2056
 Db 548 TTTTATTTAATAGTATTTGTTATTTAGTATTAATTTATTTATTTTATTTTAA 607
 QY 2057 CAATATAGTGTGGAATTCAAATATCTCCATTAATTTTGGAAATCTACAATATTT 2116
 Db 608 TATGATTTTATTTTAAATATAATTTAT-ATTTTATTTTAAATTTTATGTTTA 665
 QY 2117 AATATTTAGCAATACAAATGCAATAGAAAGTCCAAAAAATTTTGTACGAACTT 2176
 Db 666 AATATATATAGTATGATTTGTTGAGAAATGTTTAATATGAAATATG-AAAAATTAATTG 723
 QY 2177 CCAATTTTATTTTATTTATGACACAAATTAACAGATAGCAAAATTTTGTGGA 2236
 Db 724 GATTTATTTATGTTTATTTTATTTTATTTATTTATTTATTTATTTGTTTAA 783
 QY 2237 TCGAAGTATGATATACATTTACCAATTTTAAATATTAATACCTATACGCGCTCA 2296
 Db 784 ATGATATTTAATATATATATATATATTTATGCT-----TTAGGTATTAATGTTT 838
 QY 2297 AAGTATGTTATCTAGAGTGTATTAATTAATGCAATGTCGATTCAGAAATGGGCAAC 2356
 Db 839 AATTTTATTTTATTTTGAATTTTATTTTATTTATTTTATTTTATTTTGAAGTATGAT 898
 QY 2357 AATGAAACGGAATTAATTAATTTTAACTTTAAATTAATTAATTTTGAATATGCTTT 2416
 Db 899 GAAAAAAGGAAATTAATTTTATTTATTTGACCTTTGTTATTTTGAAGAAAGTAG 958
 QY 2417 TCGACTATTTGGGCAAAAAAAGACAAATGCCAAA 2454
 Db 959 GATTAAGAAAAATATGTAATAATTAATTTTGAAGA 996
 RESULT 6
 ABL32426
 ID ABL32426 standard; DNA; 9810 BP.
 AC ABL32426;
 XX
 DT 26-MAR-2002 (first entry)
 DE Human immune system associated gene SEQ ID NO: 399.
 XX
 KW Human; immune system disease; cytosine methylation; antiasthmatic;
 KW antiarteriosclerotic; anti-leukemia; cytosine methylation; antiasthmatic;
 KW neuroprotective; anti-HIV; anticonvulsant; ophthalmological;
 KW antiinflammatory; cancer; eye disease; arteriosclerosis; anaemia;
 KW acute myeloid leukemia; Alzheimer's disease; AIDS; epilepsy;
 KW neurofibromatosis; rheumatoid arthritis; psoriasis; bowel disease;
 KW gene; ds.
 XX
 OS Homo sapiens.
 XX
 PN W0200200928-A2.
 XX
 PD 03-JAN-2002.
 XX
 PF 02-JUL-2001; 2001WC-EP07537.
 XX
 PR 30-JUN-2000; 2000DE-1033529.
 XX
 PA 01-SEP-2000; 2000DE-1043826.
 XX
 PA (EPIC-) EPIGENOMICS AG.
 XX
 PI Olek A, Piepenbrock C, Berlin K.
 XX
 DR WPI; 2002-130909/17.

XX
 PT Nucleic acid comprising fragment of chemically modified gene, useful
 PT for diagnosis and treatment of diseases associated with abnormal
 PT cytosine methylation
 XX
 PS Claim 1; SEQ ID NO 399; 32pp + Sequence Listing; German.
 CC
 CC The present invention provides a number of human immune system associated
 CC genes which are modified by the methylation of cytosines. The sequences
 CC can be used in the diagnosis and treatment of immune system disorders,
 CC including eye diseases such as retinopathy, neovascular glaucoma and
 CC macular degeneration, arteriosclerosis, anaemia, cancer, acute myeloid
 CC leukemia, Alzheimer's disease, AIDS, epilepsy, neurofibromatosis,
 CC rheumatoid arthritis, psoriasis and inflammatory/ulcerative bowel
 CC diseases. The present sequence is a gene of the invention.
 SQ

Sequence 9810 BP; 2770 A; 79 C; 1978 G; 4983 T; 0 other;

Query Match 2.6%; Score 82.6; DB 24; Length 9810;
 Best Local Similarity 46.9%; Pred. No. 0.00031;
 Matches 374; Conservative 0; Mismatches 404; Indels 19; Gaps 3;

QY 1625 TAGATGATGATTTCCAAATACATTTGATGATTTAACTAATCTTTCTCT 1684
 Db 822 TTGTTGTAATGTTTGTATGATATTTGTTGTTTAACTAATCTTTCTCT 881
 QY 1685 ACGGTATTAATTAATCAATGAGTAATAA--AAGTTTGTCTATTTTGGCATGCA 1741
 Db 882 GTGAATTTTATTTATTTTGTGTTTAAATGTAATTTTATTTTATTTTATTT 941
 QY 1742 TGAAGATTAACCTTAATGATTTTATTTTATTTTATTTTATTTTATTTTATTT 941
 Db 942 TTTTATTTTATTTTATTTTATTTTATTTTATTTTATTTTATTTTATTTTATTT 1001
 QY 1802 ATTAACGTAATTTTGTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTT 1861
 Db 1002 AATATTTTATTTATTTATTTATTTATTTATTTATTTATTTATTTATTTATTT 1061
 QY 1862 CTGCAAAATTAATTAATGGAATTCATGCTACTACTACTACTACTACTACTACTACT 1921
 Db 1062 TAATATTAATGGAATAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 1121
 QY 1922 CATTTTATTTTAAACCAAAATTTTGAATA-----ATGATATATTTCT 1966
 Db 1122 TTTTATTTATTTATTTATTTATTTATTTATTTATTTATTTATTTATTTATTT 1181
 QY 1967 AACATATTTATTAATATGATGCTATTAATGTTTCTTCTTCTTCTTCTTCTTCTTCT 2026
 Db 1182 ATATATTTATTTATTTATTTATTTATTTATTTATTTATTTATTTATTTATTTATTT 1241
 QY 2027 TTTTATTTATTTATTTATTTATTTATTTATTTATTTATTTATTTATTTATTTATTT 2086
 Db 1242 ATATATTAATGTTTATTTATTTATTTATTTATTTATTTATTTATTTATTTATTT 2146
 QY 2087 ATTAATATTTTGAATCTACAAATTTATTTATTTATTTATTTATTTATTTATTTATTT 2146
 Db 1302 ATATTAATATTTATTTATTTATTTATTTATTTATTTATTTATTTATTTATTTATTT 2146
 QY 2147 TTTCAAAAAAATTTTGTACGAAATCTCCAAATTTTATTTTATTTATTTATTTATTTATTT 2205
 Db 1362 ATATTAATATTTATTTATTTATTTATTTATTTATTTATTTATTTATTTATTTATTT 1421
 QY 2206 ATTAACAGATGGAATATTTTGTGGAATGGAATGGAATGGAATGGAATGGAATGGAAT 2265
 Db 1422 AATATATTAATGTTTATTTATTTATTTATTTATTTATTTATTTATTTATTTATTTATTT 1481
 QY 2266 TTTAAAAATTTATTAAGCTTATTAAGCTTATTAAGCTTATTAAGCTTATTAAGCTTATTAAGCT 2325
 Db 1482 ATGATATTAATTAATTTGTTAAAGATATTTATTTATTTATTTATTTATTTATTTATTTATTT 1541
 QY 2326 AATGATGTTGGAATTCAGAAATTTGGGACAAATGGAATGGAATGGAATGGAATGGAATGGAAT 2385
 Db 1542 GTAAATATTAATTAATTTATTTATTTATTTATTTATTTATTTATTTATTTATTTATTTATTT 1601

ID	ABK28233 standard; DNA; 15732 BP.
XX	
AC	ABK28233;
XX	
DT	23-APR-2002 (first entry)
XX	
DE	DNA transcription associated genomic DNA #54.
XX	
KW	DNA transcription associated gene; peptide nucleic acid; PNA-oligomer;
KW	PNA; cytosine methylation state; SNP; retroviral infection; gene; ds;
KW	single nucleotide polymorphism; adenosine deaminase deficiency; cancer;
KW	viral infection; Sezary syndrome; haematological disorder; tuberculosis;
KW	immunological disorder; Werner syndrome; developmental disorder::
KW	psoriasis; Rieger's syndrome; neurological disorder; erythropoiesis;
KW	neurodegenerative disorder; Waardenburg syndrome; Niemann-Pick disease;
KW	myelodysplastic syndrome; myocardial infarction; hypertension; arthritis;
KW	angiogenesis; congenital heart disease; HDR syndrome; gene therapy;
KW	polyglutamine disorder; solid tumour.
OS	Unidentified.
XX	
PN	WO200192565-A2.
XX	
PD	06-DEC-2001.
XX	
PF	06-APR-2001; 2001WO-EP03973.
XX	
PR	06-APR-2000; 2000DE-1019058.
PR	07-APR-2000; 2000DE-1019173.
PR	30-JUN-2000; 2000DE-1032529.
PR	01-SEP-2000; 2000DE-1043826.
XX	
PA	(EPIG-) EPIGNOMICS AG.
XX	
PT	Olek A, Piepenbrock C, Berlin K;
XX	
DR	WPI; 2002-090046/12.
XX	
PT	New nucleic acids or oligomers, useful for diagnosing or treating
PT	diseases associated with DNA transcription, e.g. immunological
PT	disorders, Werner syndrome, psoriasis, myocardial infarction, solid
PT	tumours or cancer _
XX	
PS	Claim 1; SEQ ID No 107; 32pp; English.
XX	
CC	The invention relates to a nucleic acid, which comprises a segment of the
CC	chemically pretreated DNA or genes associated with DNA transcription from
CC	one of 346 sequences, and an oligomer, in particular an oligonucleotide
CC	or peptide nucleic acid (PNA)-oligomer that hybridises to or is identical
CC	to the chemically pretreated DNA of genes associated with DNA
CC	transcription. The set of oligomer probes are useful for detecting the
CC	cytosine methylation state and/or single nucleotide polymorphisms (SNPs)
CC	in a chemically pretreated genomic DNA. The nucleic acids are useful for
CC	diagnosing or treating diseases associated with DNA transcription
CC	(particularly with the methylation status), e.g. adenosine deaminase
CC	deficiency, viral infection, retroviral infection, Sezary syndrome,
CC	haematological disorders, immunological disorders, Werner syndrome,
CC	tuberculosis, developmental disorders, psoriasis, Rieger's syndrome,
CC	neurological disorders, neurodegenerative disorders, myeloid dysplasia
CC	sndrome, Niemann-Pick disease, myelodysplastic syndrome, myocardial
CC	infarction, hypertension, angiogenesis, erythropoiesis, congenital heart
CC	or cancer. Sequences ABK28127-ABK28472 represent DNA transcription
CC	associated genomic DNA molecules of the invention.
CC	Note: The sequence data for this patent did not form part of the printed
CC	specification but was obtained in electronic format directly from the
CC	European Patent Office.
XX	
SO	Sequence 15732 BP; 4638 A; 70 C; 2672 G; 8352 T; 0 other;
Query Match	2.6%; Score 82.6; DB 24; Length 15732;
Best Local Similarity	43.9%; Pred. No. 0.00032;
Matches 453; Conservative	0; Mismatches 574; Indels 6; Gaps 2;

RESULT 9
 AAS45384/c
 ID AAS45384 standard; DNA: 5689 BP.
 AC AAS45384;
 XX
 XX
 DT 18-DEC-2001 (first entry)
 XX
 DE Chemically pretreated genomic DNA associated with cell cycle #45.
 XX
 KM Cell cycle; human; Cpg dinucleotide; cytosine methylation; HIV; aging;
 KM human immunodeficiency virus; neurodegenerative disorder; solid tumour;
 KM graft-versus-host disease; glomerular disease; Lewy body disease; cancer;
 KM arthritis; arteriosclerosis; anti-HIV; neuroprotective; antiatherosclerotic;
 KM immunosuppressive; antitumour; cytostatic; antiarteriosclerotic; ds;
 KM PCR primer.
 XX
 OS Homo sapiens.
 XX
 PN WO200168911-A2.
 XX
 PD 20-SEP-2001.
 XX
 PF 15-MAR-2001; 2001WO-EP02945.
 XX
 PR 15-MAR-2000; 2000DE-1013847.
 PR 06-APR-2000; 2000DE-1019058.
 PR 07-APR-2000; 2000DE-1019173.
 PR 30-JUN-2000; 2000DE-1032529.
 PR 01-SEP-2000; 2000DE-1043826.
 XX
 PA (EPIC-) EPIGENOMICS AG.
 XX
 PI Olek A, Piepenbrock C, Berlin K;
 XX
 DR WPI: 2001-602751/68.
 XX
 PT Designing primers and probes for analysing diseases associated with
 PT cytosine methylation state e.g. arthritis, cancer, aging,
 PT arteriosclerosis comprising fragments of chemically modified genes
 PT associated with cell cycle -
 XX
 PS Claim 1; SEQ ID No 89; 28bp; English.
 XX
 CC Sequences AAS45296-AAS45520 represent chemically pretreated genomic DNA
 CC molecules associated with the cell cycle and specific PCR primers of the
 CC invention. The sequences are useful for detecting the methylation state
 CC of all Cpg dinucleotides in a sequence and therefore for analysing
 CC associated diseases. By analysing cytosine methylations in the pretreated
 CC DNA, genetic and/or epigenetic parameters for the diagnosis and therapy
 CC of existing diseases or the predisposition to specific diseases can be
 CC ascertained. The parameters may be compared to another set of genetic
 CC and/or epigenetic parameters, the differences serving as basis for
 CC diagnosis and/or prognosis events which are disadvantageous to patients.
 CC The sequences of the invention are useful for the diagnosis and therapy
 CC of HIV infection, neurodegenerative disorders, graft-versus-host disease,
 CC aging, glomerular disease, Lewy body disease, arthritis,
 CC arteriosclerosis, solid tumours and cancers.
 CC
 SO Sequence 5689 BP; 1490 A; 207 C; 1277 G; 2707 T; 8 other;

Query Match 2.5%; Score 81.6; DB 22; Length 5689;
 Best Local Similarity 49.1%; Pred. No. 0.00044;
 Matches 285; Conservative 0; Mismatches 292; Indels 3; Gaps 3;

QY 1824 ATATGACAGCCCTCTACACTGTGATGATGATTTTTCGCAATTTTAATTAGCAAT 1883
 DB 2110 AAAATGCGAATCGGCAACCGCAATATTCATAAAAAATTTCCCGCAAAAAA 2051
 QY 1884 TCAATGCTATCACTAGTAAGAAACAGCTGATTTTACATTTTAATTAAAGCAAAAT 1943
 DB 2050 TATATAAAAAAACATATAAAAAAATAAATTAATTTTATATATTTTATTTT 1991

QY 1944 TTTTGAAAAATGTTATATTTCTTACAAATATTATTAATGATGCTATATGATTT 2003
 DB 1990 ATTAATTTAAAAATATTAATAAAAAAATAAAAAATTTAATANNNTAAAAAANNTAN 1931
 QY 2004 CCTATGCTTTAAAAATATTTTATTTATATTAGTTAATAATACATTTATGAACCAATAT 2063
 DB 1930 NATAAAAAATAATATAATATAATATAATATAATATAATATAATATAATATAA 1871
 QY 2064 AGTTGGT-GAATTCAAATATCTCCATTAATATTTTGGAAATCTCAATATTTAATAT 2122
 DB 1870 AATAAATATTTAATTAATTAATTTATTTATATATTTAAATTAATTAATTAATTA 1811
 QY 2123 TAGTCATTAACATGATCAT-AGAAAGTCCAAAAAATTTGTTAACGAAACTTCCAA 2181
 DB 1810 TATATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAA 1751
 QY 2182 TTTTATTTTATTAAGAACAAAGAAATTAACAGATAGAAACATTTTGTGGAAATGGA 2241
 DB 1750 TAAATATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAA 1691
 QY 2242 GTAGTAATATATCATTAAGCAATTTTAAATAATTAATAGCTATACGCGCTCAAGTA 2301
 DB 1690 TAACAATTAATAAATAATACGAATATTTAATTTCAATTAATTAATAAACAACAAACAA 1631
 QY 2302 TGTATATAGTAGGTGTAATTAATA-ATGCATGTGCGATTCCAGATTGGGCAACAATG 2360
 DB 1630 TATATATCATCAAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAA 1571
 QY 2361 AAAACGAAATTAATAATATTAATCTTTAAATAAATAAATAAAT 2400
 DB 1570 ATAAAAAATAATTTAATTAACGTTAAATAAATAAATAATAT 1531

RESULT 10
 AAS46426/c
 ID AAS46426 standard; DNA: 5689 BP.
 AC AAS46426;
 XX
 XX
 DT 18-DEC-2001 (first entry)
 XX
 DE Tumour suppressor gene derived chemically modified sequence #148.
 XX
 KM Human; tumour suppressor gene; oncogene; antitumour; cytostatic;
 KM cancer; tumour; Cpg dinucleotide; single-nucleotide polymorphism; SNP;
 KM cytosine methylation; ds.
 XX
 OS Homo sapiens.
 XX
 PN WO200168912-A2.
 XX
 PD 20-SEP-2001.
 XX
 PF 15-MAR-2001; 2001WO-EP02955.
 XX
 PR 15-MAR-2000; 2000DE-1013847.
 PR 06-APR-2000; 2000DE-1019058.
 PR 07-APR-2000; 2000DE-1019173.
 PR 30-JUN-2000; 2000DE-1032529.
 PR 01-SEP-2000; 2000DE-1043826.
 XX
 PA (EPIC-) EPIGENOMICS AG.
 XX
 PI Olek A, Piepenbrock C, Berlin K;
 XX
 DR WPI: 2001-602752/68.
 XX
 PT Fragments of chemically modified genes associated with tumour suppressor
 PT genes and oncogenes, useful in designing primers and probes for
 PT analysing diseases associated with cytosine methylation state e.g.
 PT cancer -
 XX

The invention relates to a nucleic acid comprising a sequence of 18 bases, of a segment of chemically pretreated DNA (CP DNA) e.g. with bisulphite, of genes associated with tumour suppression and oncogenes having a sequence taken from 536 (actually 533 since numbers 408, 458 and 500 are missing from the sequence listing) sequences (Ss) and sequences complementary to (Ss). The nucleic acid may be a peptide nucleic acid-oligomer (PNA) of at least 9 nucleotides and may form part of a set of probes for detecting the cytosine methylation state and/or single nucleotide polymorphisms and also to be used in an array for analysing diseases associated with CpG dinucleotides e.g. cancers and tumours. The probes can also be used in a method for ascertaining genetic and/or epigenetic parameters for the diagnosis and/or therapy of existing diseases or the predisposition to specific diseases, by analysing cytosine methylations. The parameters may be compared to another set of genetic and/or epigenetic parameters, the differences serving as basis for diagnosis and/or prognosis events which are disadvantageous to patients. The present sequence is one of the 533 genomic sequences derived from tumour suppressor genes and oncogenes. Sequences with even numbered Seq ID numbers are the complementary sequence of the corresponding odd numbered sequence (e.g. ID 2 and ID1, ID 536 and ID 535, except for those whose partner sequence is missing).

Note: The sequence data for this patent did not form part of the printed specification, but was obtained in electronic format directly from WIPO at ftp.wipo.int/pub/published_pct_sequences.

50 Sequence 5689 BP; 1490 A; 207 C; 1277 G; 2707 T; 8 other;

Query Match	2.58;	Score 81.6;	DB 22;	Length 5689;
Best Local Similarity	49.1%;			
Matches 285;	Conservative	0.00044;		
	Mismatch			

Qy	1824	ATATGACAGCCCTTCACAACTGATAGTCAATTTTTCGAAATATTTAAATAGGAAT	1883
Db	2110	AAATCGCAATCCGCAACCGGCAATATATCAAAAAATTTCTCCGAAAAA	2051
Qy	1884	TCATGCTACTATCATAGACAGAACAGCTGAGTATTCATTTTATTTAAAGACAAAT	1943
Db	2050	TATAAAAAACAATTAATAAATAATTTATTTTATATATTTTTTTTTTT	1991
Qy	1944	TTTGGAAAAATGTATTAATTTCAACAATATATTAATAAATAGAGCTATATATGATTT	2003
Db	1990	ATTATTTAAAAATTAATTAATAAATAAATAAATTTATATANNTTAAAAAANNNTAN	1931
Qy	2004	CCATGCTCTTAAATATTTTATTTTATATTTAGTTATTAATATACATATGACCAATAT	2063
Db	1930	NATAAAAAATAATTAATTAATTAATAAATAAATTTATTAATAAATAAATAAATAA	1871
Qy	2064	ACTTGGT-GAATCCAATATCTCATTAATTTTGGAAATCTCAAAATATTAATATT	2122
Db	1870	AATAAATATTTATTAATAATTTATTTATATTAATTAATAAATAAATTAATTAATAAAT	1811
Qy	2123	TAGTCATAATACAAATGCAAT-AGAAAGTCCAAAAAATTTGTTAACAGAACTCCAAAT	2181
Db	1810	TATAAAAATAAAAATTAATTAATAAATAATTAATAAATAAATAAATAAATAATATATAT	1751
Qy	2182	TTTTTTTTTTTATGGAACAAGAAATTAACGATAGAAAACTATTTTGTGTGGAATGAA	2241
Db	1750	TAAATTAATAATTAATAAATAAATAAATAAATAATTAATAAATAAATAAATAAATAA	1691
Qy	2242	GTAATATATATACATTAAGCAAAATTTTAAAAAATTAATATTAAGCTATACGCGTCAAGTA	2301
Db	1690	TAACAATTAATAATTAAGCAATATATTAATTAATTAATTAATTAATTAATTAATTAAT	1631
Qy	2302	TGTTATCTACTAGCTGTAATTAATA-ATGCAATGCTCGGATGAGATTTGGGCAACAATG	2360
Db	1630	TATTAATCTACAAAAATTAATAAATAATTAATTAATTAATTAATTAATTAATTAATTAAT	1571
Qy	2361	AAAACGAATTAATAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT	2400

Db 1570 ATAAAAAAAAATTTTAATACGTTAAAAATATAAATATAT 1531

RESULT 11

ABK28226/c
ID ABK28226 standard; DNA; 5689 bp.
XX

AC ABK28226;

D1 23-APR-2002 (first entry)
XX

transcription associated complementary genomic DNA #50.

KW pNA: cyano methylated gene; peptide nucleic acid; PNA-oligomer;
 KW single nucleotide polymorphism; SNP; retroviral infection; gene; ds;
 KW viral infection; leishmaniasis; adenosine deaminase deficiency; cancer;
 KW immunological disorder; haematological disorder; tuberculosis;
 KW porosis; Reiger's syndrome; neurological developmental disorder;
 KW neurodegenerative disorder; Waardenburg syndrome; Meeman-Pick disease;
 KW myelodysplastic syndrome; myocardial infarction; hypertension; arthritis
 KW polyglutamine disorder; congenital heart disease; HDR syndrome; gene therapy;
 KW solid tumour.

Unidentified

WO200192565-A2

06-DEC-2001.

06-APR-2001; 2001WO-EP03973

00-APR-2000; 2000DE-1019058
07-APR-2000; 2000DE-1019173

01-SEP-2000; 2000DE-1043826

(EPiG-) EPIGENOMICS AG.

Olek A, Piepenbrock C, Berlin K;

WPI; 2002-090046/12.

new nucleic acids or oligomers, useful for diagnosing or treating diseases associated with DNA transcription, e.g. immunological disorders, Werner syndrome, psoriasis, myocardial infarction, solid tumours or cancer.

Claim 1; SEQ ID No 100; 32pp; English.

chemically-pretreated DNA of genes associated with a nucleic acid, which comprises a segment of the one of 346 sequences, and an oligomer, in particular an oligonucleotide or peptide nucleic acid (PNA)-oligomer that hybridises to or is identical to the chemically pretreated DNA of genes associated with DNA transcription. The set of oligomer probes are useful for detecting the cytosine methylation state and/or single nucleotide polymorphisms (SNPs) in a chemically pretreated genomic DNA. The nucleic acids are useful for diagnosing or treating diseases associated with DNA transcription (particularly with the methylation status), e.g. adenoma desmase deficiency, viral infection, retroviral infection, Sezary syndrome, haematological disorders, immunological disorders, Werner syndrome, tuberculosis, developmental disorders, psoriasis, Rieger's syndrome, neurological disorders, neurodegenerative disorders, Wardenburg syndrome, Niemann-Pick disease, angiodysplasia, congenital heart infection, hyperextension, anxiogenesis, erythropoiesis, myocardial disease, HNR syndrome, arthritis, polyglutamine disorders, solid tumours or cancer. Sequences ABK28127-ABK28472 represent DNA transcription associated genomic DNA molecules of the invention.

Note: The sequence data for this patent did not form part of the printed specification but was obtained in electronic format directly from the European Patent Office.

RESULT 14

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: June 10, 2003, 22:02:58 ; Search time 152.509 Seconds

(without alignments)
6438.831 Million cell updates/sec

Title: US-09-502-426B-1_COPY_1_3202

Perfect score: 3202

Sequence: 1 atgtgggtattatattgttg.....gagagagagaactactctcc 3202

Scoring table: IDENTITY_NUC

Gapop 10.0 , Gapext 1.0

Searched: 441362 seqs, 15338381 residues

Total number of hits satisfying chosen parameters: 882724

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 10%

Listing first 45 summaries

Database : Issued_Patents_NA.*
1: /cg2.6/ptodata/1/ina/5A_COMB.seq:*
2: /cg2.6/ptodata/1/ina/5B_COMB.seq:*
3: /cg2.6/ptodata/1/ina/6A_COMB.seq:*
4: /cg2.6/ptodata/1/ina/6B_COMB.seq:*
5: /cg2.6/ptodata/1/ina/PCUTS_COMB.seq:*
6: /cg2.6/ptodata/1/ina/Backfills1.seq:*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result	No.	Score	Query Match	Length	DB ID	Description
1	77.4	2.4	19124	2	US-08-487-826B-13	Sequence 13, Appl
2	68.6	2.1	837	4	US-08-998-416-288	Sequence 288, App
3	67.4	2.1	19124	2	US-08-487-826B-13	Sequence 13, Appl
4	65.8	2.1	615	4	US-08-998-416-186	Sequence 186, App
5	65.4	2.0	636	4	US-08-998-416-1137	Sequence 1137, App
6	65.4	2.0	20674	4	US-09-641-638-651	Sequence 651, App
7	63.6	2.0	8920	2	US-08-446-855A-1	Sequence 1, Appl
8	63.6	2.0	8920	2	US-09-150-741-1	Sequence 1, Appl
9	63.6	2.0	20674	4	US-09-641-638-651	Sequence 651, App
10	62.6	2.0	4818	3	US-08-817-926-27	Sequence 27, Appl
11	61.8	1.9	5852	1	US-07-867-106-2	Sequence 2, Appl
12	61.2	1.9	1431	4	US-09-316-083-2	Sequence 2, Appl
13	60.8	1.9	4818	3	US-08-817-926-27	Sequence 27, Appl
14	60.6	1.9	837	4	US-08-998-416-288	Sequence 288, App
15	60.4	1.9	8920	2	US-08-446-855A-1	Sequence 1, Appl
16	60.4	1.9	8920	2	US-09-150-741-1	Sequence 1, Appl
17	60.2	1.9	665	2	US-08-883-795A-36	Sequence 36, Appl
18	60	1.9	1864	4	US-09-468-265-4	Sequence 4, Appl
19	59.8	1.9	6124	4	US-08-213-419B-3	Sequence 3, Appl
20	59.2	1.8	51952	4	US-08-947-823-1	Sequence 1, Appl
21	59	1.8	827	4	US-08-998-416-535	Sequence 535, App
22	58.4	1.8	665	2	US-08-883-795A-36	Sequence 36, App
23	58.4	1.8	12793	4	US-09-004-838-124	Sequence 124, App
24	58.2	1.8	6124	4	US-08-213-419B-3	Sequence 3, Appl
25	57.8	1.8	1511	1	US-07-991-867B-8	Sequence 8, Appl
26	57.8	1.8	1511	1	US-08-107-755A-8	Sequence 8, Appl
27	57.8	1.8	1511	2	US-08-544-332-8	Sequence 8, Appl

c	28	57.8	1.8	1511	4	US-09-370-861A-8	Sequence 8, Appl
	29	57.4	1.8	3095	6	US-09-426-290-1	Patent No. 5231168
	30	56.6	1.8	168575	4	US-08-749-391-1	Sequence 1, Appl
c	31	56.4	1.8	2058	2	US-07-638-431-1	Sequence 1, Appl
	32	56.4	1.8	2058	2	US-09-390-200-1	Sequence 1, Appl
c	33	56.4	1.8	4673	3	US-07-638-431-1	Sequence 1, Appl
	34	56.4	1.8	4673	5	PCT-US92-00018-1	Sequence 2, Appl
c	35	56.2	1.8	1431	4	US-09-316-083-2	Sequence 1, Appl
	36	56.2	1.8	2960	3	US-08-913-842-3	Sequence 3, Appl
c	37	55.6	1.7	701	4	US-08-998-416-701	Sequence 701, App
	38	55.6	1.7	5852	1	US-07-867-106-2	Sequence 2, Appl
c	39	55.2	1.7	1406	3	US-08-913-842-6	Sequence 6, Appl
	40	55.2	1.7	1511	1	US-07-991-867B-8	Sequence 8, Appl
c	41	55.2	1.7	1511	1	US-08-107-755A-8	Sequence 8, Appl
	42	55.2	1.7	1511	2	US-08-544-332-8	Sequence 8, Appl
c	43	55.2	1.7	1511	4	US-09-370-861A-8	Sequence 8, Appl
	44	55	1.7	1850	4	US-08-617-860B-32	Sequence 32, Appl
c	45	55	1.7	4098	2	US-08-605-106-4	Sequence 4, Appl

ALIGNMENTS

RESULT 1
US-08-487-826B-13
; Sequence 13, Application US/08487826B
; Patent No. 5938627
; GENERAL INFORMATION:
; APPLICANT: Sim, Kim L.
; APPLICANT: Chitnis, Chetan
; APPLICANT: Miller, Louis H.
; APPLICANT: Peterson, David S.
; APPLICANT: Su, Xin-zhaun
; APPLICANT: Wellens, Thomas E.
; TITLE OF INVENTION: BINDING DOMAINS FROM PLASMODIUM VIVAX
; NUMBER OF SEQUENCES: 45
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Knodde Martens Olson & Bear
; STREET: 620 Newport Center Drive 16th Floor
; CITY: Newport Beach
; STATE: California
; COUNTRY: US
; ZIP: 92660
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/487,826B
; FILING DATE: 10-SEP-1993
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Israelsen, Ned
; REGISTRATION NUMBER: 29,655
; REFERENCE/DOCKET NUMBER: NIH121.001CPI
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (619) 235-8550
; TELEFAX: (619) 235-0176
; INFORMATION FOR SEQ ID NO: 13:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 19124 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: cDNA
; HYPOTHETICAL: NO
; ANTI-SENSE: NO
; US-08-487-826B-13
Query Match 2.4%; Score 77.4; DB 2; Length 19124;
Best Local Similarity 42.6%; Pred. No. 1.8e-06;

	Matches	590:	Conservative	0:	Mismatches	781:	Indels	14:	Gaps	3:
OY	1013	TTAAATACGATTTTACAGCGTCACTAGATTACTAGCATTAAGCATTAAGAGACC	1071Z							
Db	229	TTATATATCAATATATATATATAGTTCCCTTTAAATTAATATATATATCAATATAT	288							
OY	1073	GTTCAAGCTATTTATACAAAGTTTCAACCTGAATATAGCTTGAATCCTTAGAAATTT	1132Z							
Db	289	TTCTTGTTATTTTATTAATATTAACCTAATTTCTTATTTTATTTAATCACTTTATTCCTTTT	348							
OY	1133	TGGAATPACCGGTGTATGTAAATATAGATTTAGTGTAAACAAATGTTAATCAAT	1192Z							
Db	349	AATTTCTTAATCTTTTATGCAACCAAAAACATTAAGTATTTCTACTATTCACACAAAA	408							
OY	1193	AGTGTCAACATTAATCATATATCTCTACAGAAAAACAACTTAAGAGAGTTAACAT	1252Z							
Db	409	AAAAAAAAAAAAAAAAAAAAATTTATTAATATATATAAATAATATAAGACATACGT	468							
OY	1253	CCATATATGGGTATGCTATACCTTCACGTATGCTATGCTATGACACTTAACATATGTT	1312Z							
Db	469	TCACCTATTTATTAATATGATTTATTCAGATTTAAACCTTATGAGATTATATATATAA	528							
OY	1313	GTGATGTGATTAATGAAATTCACACGCGTGTATATATATTTGGAGCGCTGTATGAT	1372Z							
Db	529	TTTAAACATAGAAAGGTAAAGATACATTTTTTTTTTTTTTTTGATATGTAATTCACAT	588							
OY	1373	CACGTCAAAATTCATCTTGTGTGTCACATATAAACAACAAAAACAGAAAAAGAAA	1432Z							
Db	589	ATATATATATATATATCTTTTATTTATTTAAATTAATTAATCTTATATATCAATGTG	648							
OY	1433	CGATTTTCTTGATTCATTCATGATGATCAATTAATGCAATATCTTTGGTTCACGTT	1492Z							
Db	649	TCTTTTATC-----ACATGTGAATATTAATAAATTAATTTGCAATTTTATTCGATATAT	702							
OY	1493	CGAAGTCCCTACAGCGTGTACACATCTGCACATTTAAATTCCTTCTTAAATGCATC	1552Z							
Db	703	ATGTCGTTATATACTTATATATAGGCTTTATTAACATATGATTAATAGAGTATATAGCT	762							
OY	1553	TTTAACATATTTATTTAGTTGGAATTTAATAAGACGAACTGTACATTAACAATTT	1612Z							
Db	763	AATATATATAAATCTCGATTTATTAATAATCTTTATATATTTCAAAATATATTCGATG	822							
OY	1613	TATATTAAGATCTGTATGTATTTATTCCAATTCATACCTTGGATGTTAACTATATC	1672Z							
Db	823	TTATTTTCAATTAACAATTAATATAGATTTCTTAAATATTTCTCTCAATTAATCATTTTATA	882							
OY	1673	TTGTTTCTCCACGATTAATAATTTATATCATTCAGGTAAAAAAAGTTTGTCTATATTT	1732Z							
Db	883	GCATATATCATGCACATTAATAATTAATTAATTAATTAATTAATTAATTAATTAATAA	942							
OY	1733	CGCATGCAATGAAAGATTAACCTTAATGACTTTAATTTTGGAAAAATGTAAACCTTTTACT	1792Z							
Db	943	ATTTTCATACATTCATTTTTCACACAACTTTAAGTGTGCATATATTAACACATTAAT	1002Z							
OY	1793	CATAGATTAATTAACCGTATGTTTTTGTGGCATATATGACGCCCTACACTGTGATAGT	1852Z							
Db	1003	AATATTAATCTTAATATATATTAATTAATTTATTAATTAATTAATTAATTAATTAATTAAT	1062Z							
OY	1853	CAATTTTCTGCAAAATTAATTAATTAAGCAATTCATGCTACTATCATTAAGAGAACAGC	1912Z							
Db	1063	CTGATTAATCATATATATATATATAGTGTATTAACGCTTCAAAATATATATATATATTA	1122Z							
OY	1913	TGACTATTAATTTAAATTAAGACAAAAATTTTGGAAAAATGTTAATATTTTCAACAT	1972Z							
Db	1123	TTAAAAATATATATATAGTAAATTAATTAATTTTGTATGTTATGAATTAATGAATTAATA	1182Z							
OY	1973	ATTTTAAATATGATGCTATATATGATTTTCCATGTCTTTAAATATTTTTTTTATA	2032Z							
Db	1183	TAAATATAAATCTATAGCTATTTTAAAAATATATAT-----ATATATATATATATATA	1237Z							
OY	2033	TTTGATTAATAATCAATTAAGAACCAATTAATAGTGTGATGCTCAATATATTCATTAAT	2092Z							
Db	1238	TATATATTAATAGTATATGTTATCAAAATATTAATATATTAATTT-----ATATATAAATAATA	1294Z							

Oy	ATTTTGGAAATTACAAAATTATTAATTTTTAGTCGAATAACAATGCAATAGAAAAGTCCAA	21522
Dd	TTTGTAATACCTAACAGACTAAGAACACTATACAACTCGTATCTTAATAGTATATATAT	13544
Oy	AAAAAATTTTGTTAAACAGAAACCTCCAAATTTTTTTTTTTTATGGAACAAAGAAATACAG	22122
Dd	AATAATCCTTTTTTATTTATTTAAATTTGCTCTCTCTTTTTTTTTTTTAAATAATATATATAT	14144
Oy	ATGAGAAAATATTTTTGTTGGAAATGCAAGTAGTAAATATACATTATACCAATTTTTAAAAA	22722
Dd	TAAATATATTTTTTTTCCATAATATATATGATTTAGTATTTTAAATATAATAATAATCTTTTAA	14744
Oy	ATTATATATAGCCTATACGCCTCAAAGATATCTTCTAGTAGGTGTAATTAATATGCAT	23222
Dd	AAAACTCCAAACACATTTTTTGCTAAATAATTAATTAATTTGTAACCCTGATGTAAATTT	15344
Oy	GGTGCGATTCGAATTTGGGACAACAATGAAGAAGAAATTAATAATTAACCTTAAAAATTA	23922
Dd	AGAGAAGAAAGCTAACAACATTAACCAAAAAAAAAAATTAAGACAAAAAGAATATTACAAAAAATAA	15944
Oy	ATTAATTA	2393
Dd	TAAAA TAAAA	1599

RESULT 2
 US-08-998-416-288
 : Sequence 288, Application US/08998416
 : Patent No. 6239264
 : GENERAL INFORMATION:
 : APPLICANT: Philippsen, Peter
 : APPLICANT: Pohlmann, Rainer
 : APPLICANT: Steiner, Sabine
 : APPLICANT: Mohr, Christine
 : APPLICANT: Wendland, Jürgen
 : APPLICANT: Knechtle, Philipp
 : APPLICANT: Redischung, Corinne
 : TITLE OF INVENTION: GENOMIC DNA SEQUENCES OF ASHBYA GOSSPYIIT
 : TITLE OF INVENTION: AND US\$ THEREOF
 : NUMBER OF SEQUENCES: 1152
 : CORRESPONDENCE ADDRESS:
 : ADDRESSEE: No. 6239264artis Corporation
 : STREET: 3054 Cornwalls Road
 : CITY: Research Triangle Park
 : STATE: No. 6239264th Carolina
 : COUNTRY: USA
 : ZIP: 27709
 : COMPUTER READABLE FORM:
 : MEDIUM TYPE: floppy disk
 : COMPUTER: IBM PC compatible
 : OPERATING SYSTEM: PC-DOS/MS-DOS
 : SOFTWARE: Patentin Release #1.0, Version #1.30
 : CURRENT APPLICATION DATA:
 : APPLICATION NUMBER: US/08/998,416
 : FILING DATE: 24-DEC-1997
 : CLASSIFICATION: 435
 : PRIOR APPLICATION DATA:
 : APPLICATION NUMBER: CH 0016/97
 : FILING DATE: 31-DEC-1996
 : ATTORNEY/AGENT INFORMATION:
 : NAME: Meigs, J. Timothy
 : REGISTRATION NUMBER: 38, 241
 : REFERENCE/DOCKET NUMBER: PF/5-30306/A/CGC1976
 : TELECOMMUNICATION INFORMATION:
 : TELEPHONE: 919-541-8587
 : TELEFAX: 919-541-8689
 : INFORMATION FOR SEQ ID NO: 288:
 : SEQUENCE CHARACTERISTICS:
 : LENGTH: 837 base pairs
 : TYPE: nucleic acid
 : STRANDEDNESS: single
 : TOPOLOGY: linear

MOLECULE TYPE: DNA (genomic)
ORIGINAL SOURCE:
ORGANISM: PAG1241RP
US-08-998-416-288

Query Match 2.1%; Score 68.6; DB 4; Length 837;
Best Local Similarity 46.9%; Pred. No. 4.6e-05;
Matches 250; Conservative 0; Mismatches 279; Indels 4; Gaps 1;

```

OY 1750 AACCTAATGACCTTAATTTTGAAGAAATGAACCCCTTACTCATGATTAAATACCGT 1809
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 56 ATATTTAAGTATTAATTTATTTAAACATATTTATTTATTTAAATTAATTAATTTATTTG 115
OY 1810 ATGTTTGTGTCATATGACACCCCTTACACTGATGATGATGATTTTTCGCAAT 1669
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 116 ATATTAATTAATCTTATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 175
OY 1870 ATTAATTAAGAAATTCATGCTACTATCAATAGAGAACAAGCTGACTATTAATTTAA 1929
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 176 TATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 235
OY 1930 TTTAAAGACAAATTTTGAAGAAATGTTATATTTCAACAATTTATTAATTAATGATG 1989
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 236 ATTAGTATTAATAAATATTAATTAATTAATTAATTTCTTAAATTAATTAATTAATGAT 295
OY 1990 CCTAATATGATTTCCATGCTCTTAATAATTTT---TTTAAATTTAGTTATTAAT 2045
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 296 ATCAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 355
OY 2046 ACATTAATGAACATATAGTGTGATGATCAATTCATCAATTAATTTTGAATC 2105
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 356 AAGTTTAATTTATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 415
OY 2106 TACAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 2165
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 416 AATATTAATGATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 475
OY 2166 AACAGAACTTCCAAATTTTATTTTGAAGAAATTAATTAATTAATTAATTAATTAAT 2225
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 476 TAACCTTAATTTCTTATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 535
OY 2226 TGTGTTGAATGGAAGTATATATATTAATTAATTAATTAATTAATTAATTAATTAAT 2278
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 536 TATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 588

```

RESULT 3

US-08-487-826B-13/C
Sequence 13, Application US/08487826B
Patent No. 5993827

GENERAL INFORMATION:
APPLICANT: Sim, Kim L.
APPLICANT: Chiltons, Chetan
APPLICANT: Miller, Louis H.
APPLICANT: Peterson, David S.
APPLICANT: Su, Xin-zhuan
APPLICANT: Wellens, Thomas E.
TITLE OF INVENTION: BINDING DOMAINS FROM PLASMODIUM VIVAX
TITLE OF INVENTION: AND PLASMODIUM FALCIPARUM ERYTHROCYTE BINDING PROTEINS
NUMBER OF SEQUENCES: 45
CORRESPONDENCE ADDRESS:
ADDRESSEE: Knodde Martens Olson & Bear
STREET: 620 Newport Center Drive 16th Floor
CITY: Newport Beach
STATE: California
COUNTRY: US

ZIP: 92660
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent Release #1.0, Version #1.25
CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/08/487,826B
FILING DATE: 10-SEP-1993
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: Israelson, Ned
REGISTRATION NUMBER: 29,655
REFERENCE/DOCKET NUMBER: NIH121.001CPI
TELECOMMUNICATION INFORMATION:
TELEPHONE: (619) 235-8550
TELEFAX: (619) 235-0176
INFORMATION FOR SEQ ID NO: 13:
SEQUENCE CHARACTERISTICS:
LENGTH: 19124 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: cDNA
HYPOHETICAL: NO
ANTI-SENSE: NO
US-08-487-826B-13

Query Match 2.1%; Score 67.4; DB 2; Length 19124;
Best Local Similarity 42.3%; Pred. No. 0.00016;
Matches 821; Conservative 0; Mismatches 1091; Indels 29; Gaps 7;

```

OY 420 TAGGTTTATTTGATCATATTAATTAATGATCATATTTCTTGATTAATTAATTAATTAAT 479
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 7044 TATGATTAATGTTAAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 6985
OY 480 CCAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 539
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 6984 TATTTTAAAAAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 6925
OY 540 CATTATGATGACTGCAATCATGATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 599
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 6924 CAAAAAATAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 6885
OY 600 AAGATTTGAGAAAAAACCAG---AATAGAAAAAGGAAGAGTGTGACCCATGAG 656
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 6864 TATATTAATTAATTAATTTTATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 6805
OY 657 TATGTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 716
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 6804 AAAAAATGCAAAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 6745
OY 717 CCGCGCAAGCTTCTGCAATCAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 776
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 6744 TATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 6685
OY 777 AAAAAATGCAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 836
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 6684 TATATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 6625
OY 837 TCCAGGTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 896
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 6624 ATCAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 6565
OY 897 CTCTTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 956
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 6564 TATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 6505
OY 957 TGGGACATTTGAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 1016
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 6504 TTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 6445
OY 1017 AATAGATTTTACAGGCTGACTGATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 1076
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 6444 ATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 6388
OY 1077 AACCTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 1136
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 6387 TATTTTAAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 6328

```

QY 1137 ATTACCGCTTGTATGTAATATAGATTGTCGTGAACAAATATGTTAATCAATAGTG 1196
DB 6327 AAATTAAGTAAAAATACACATTAATAATATATATCAAAATATGAGTTAATAATATG 6268
QY 1197 GTCAACATATACATTAATTCCTTACAGAAAAAACAACCTTAAGAGAACTTAACATATCCAT 1256
DB 6267 TTATATGTCATATATATTTATATTAAGAAAAATTTGTTATATATTAATTAATATATGTC 6208
QY 1257 ATATGGGTATGCTATACCTTTTAC -GTATGCTATAGTACAGACTTAAGAACTATATGTC 1315
DB 6207 TACTATATTAATATTAATATATATCTTAAGTATATATCAATAATATATTAATATGATGT 6148
QY 1316 ATGTCCATTAATGAATTCACACGCGGTGTAATATATGAGACCGTATGTCATGCATC 1375
DB 6147 ATAAAAATAGTAAATATCAATATATATATATATATATATATATATATATATATATATA 6088
QY 1376 TGCATAATATCATCTTGGTGTGCAACATTAATAAACAACAAAAACAAGAAAAACGA 1435
DB 6087 TATATATATATATGATATATATGTAATTAATTTATTTATATTCCAATATCTGATATGTT 6028
QY 1436 TTTTCTTGATTCATTCATATGATATATTAATGATATATGATATGATATGATATGAT 1495
DB 6027 TTTATATTTGTTA 5968
QY 1496 AGTCCCTACACACGCTGTAACCATCTG -CACTATTAATTTGCTTTCTTAATGATCT 1553
DB 5967 AT 5908
QY 1554 TTAACAT 1613
DB 5907 TAT 5848
QY 1614 ATATGAT 1673
DB 5847 AGTAT 5788
QY 1674 TGTTCCTCTACGAT 1733
DB 5787 GTAT 5728
QY 1734 GCGATGATGATGAT 1793
DB 5727 TGTATATCA -CAT 5669
QY 1794 ATGAT 1853
DB 5668 CCAATCAAAAAACAT 5609
QY 1854 AATTTTCTGCAAAAT 1913
DB 5608 TATTTTGTATGAATATCTAT 5559
QY 1914 GAGTAT 1973
DB 5558 ---TATTTAAATTTTTCAAAAAATAATCATGTTAATATATATATATATATATATATAT 5502
QY 1974 TAT 2033
DB 5501 TGACAT 5442
QY 2034 TTAGTAT 2090
DB 5441 TTTCTAT 5382
QY 2091 ---ATATTTTGAAT 2147
DB 5381 GACTTAT 5322
QY 2148 TCCAAAAAATTTTGTATACAGAACTTCAAAATTTTGTATATATATATATATATATATAT 2207
DB 5321 TATTTACTTAT 5262
QY 2208 AACAGAT 2267

DB 5261 TAAGCTATATGCACTACAAATCAGAAATGATATATATATATATATATATATATATATAT 5202
QY 2268 AAAAAAT 2327
DB 5201 AAT 5142
QY 2328 TGCATGCTGAT 2348
DB 5141 AAAAAACAAATATGAAAT 5121

RESULT 4.
US-08-998-416-186
Sequence 186, Application US/08998416
Patent No. 6239264
GENERAL INFORMATION:
APPLICANT: Philippsen, Peter
APPLICANT: Pohlmann, Rainer
APPLICANT: Steiner, Sabine
APPLICANT: Mohr, Christine
APPLICANT: Wendland, Jurgen
APPLICANT: Knechtel, Philipp
APPLICANT: Redischung, Corinne
TITLE OF INVENTION: GENOMIC DNA SEQUENCES OF ASHBYA GOSSYPIT
TITLE OF INVENTION: AND USES THEREOF
NUMBER OF SEQUENCES: 1152
CORRESPONDENCE ADDRESS:
ADDRESSEE: No. 6239264artis Corporation
STREET: 3054 Cornwallis Road
CITY: Research Triangle Park
STATE: No. 6239264th Carolina
COUNTRY: USA
ZIP: 27709
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/998,416
FILING DATE: 24-DEC-1997
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: CH 0016/97
FILING DATE: 31-DEC-1996
ATTORNEY/AGENT INFORMATION:
NAME: Meigs, J. Timothy
REGISTRATION NUMBER: 36,241
REFERENCE/DOCKET NUMBER: PF/5-30306/A/CGC1976
TELECOMMUNICATION INFORMATION:
TELEPHONE: 919-541-8587
TELEFAX: 919-541-8689
INFORMATION FOR SEQ ID NO: 186:
SEQUENCE CHARACTERISTICS:
LENGTH: 615 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: DNA (genomic)
ORIGINAL SOURCE:
ORGANISM: PAG107ARP
US-08-998-416-186

Query Match 2.1%; Score 65.8; DB 4; Length 615;
Best Local Similarity 48.3%; Pred. No. 0.00015;
Matches 216; Conservative 0; Mismatches 227; Indels 4; Gaps 1;

QY 1750 AACCTAATGACTTATATTTTGAATATTAACCTTTTACTATAGATTAATTAACCT 1809
DB 56 ATATTTAAGATTAATATATATATATTAACATATATATATATATATATATATATATAT 115
QY 1810 ATGTGTTTGTGCAATATGACAGCCTTAACAACCTGATATGCAATTTTCTGCAAT 1869

[illegible]

```

RESULT 5
US-08-998-416-1137
: Sequence 1137, Application US/08998416
: Patent No. 6239264
:
: GENERAL INFORMATION:
: APPLICANT: Philippsen, Peter
: APPLICANT: Pohlmann, Rainer
: APPLICANT: Steiner, Sabine
: APPLICANT: Mohr, Christine
: APPLICANT: Wendland, Jürgen
: APPLICANT: Knechtel, Philipp
: APPLICANT: Reblsching, Corinne
: TITLE OF INVENTION: GENOMIC DNA SEQUENCES OF ASHBYA GOSYPHYLLIDAE
: TITLE OF INVENTION: AND USUS THEREOF
: NUMBER OF SEQUENCES: 1152
:
: CORRESPONDENCE ADDRESS:
: ADDRESSEE: No. 6239264artis Corporation
: STREET: 3054 Cornwallis Road
: CITY: Research Triangle Park
: STATE: No. 6239264th Carolina
: COUNTRY: USA
: ZIP: 27709
:
: COMPUTER READABLE FORM:
: MEDIUM TYPE: Floppy disk
: COMPUTER: IBM PC compatible
: OPERATING SYSTEM: PC-DOS/MS-DOS
: SOFTWARE: PatentIn Release #1.0, Version #1.30
: CURRENT APPLICATION DATA:
: APPLICATION NUMBER: US/08/998,416
: FILING DATE: 24-DEC-1997
: CLASSIFICATION: 435
:
: PRIOR APPLICATION DATA:
: APPLICATION NUMBER: CH 0016/97
: FILING DATE: 31-DEC-1996
: ATTORNEY/AGENT INFORMATION:
: NAME: Meigs, J. Timothy
: REGISTRATION NUMBER: 38,241
: REFERENCE/DOCKET NUMBER: PF/5-30306/A/CGC1976
: TELECOMMUNICATION INFORMATION:
: TELEPHONE: 919-541-8587
: TELEFAX: 919-541-8689
: INFORMATION FOR SEQ ID NO: 1137:
: SEQUENCE CHARACTERISTICS:
: LENGTH: 636 base pairs
: TYPE: nucleic acid

```

```

; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
; ORIGINAL SOURCE:
; ORGANISM: PAG1692RP
;
US-08-998-416-1137

```

Query Match	2.0%;	Score 65.4;	DB 4;	Length 636;
Best Local Similarity	47.1%;	Pred. No. 0.00018;		
Matches 269;	Conservative 0;	Mismatches 296;	Indels 6;	Gaps 2;

QY	1341	AACGCTAATCATTTTCTTCGCAATATTTAAATTTAGAAATTCAGTCAATGCTATCAT	1900
Db	29	AAGATTATATAAACCTTTTTTATATATATATTTAAGTATTAATTTAAACCTATAT	88
QY	1901	AGAAGAAACAGCTGAGTATTCATTTTATTTAAAGACAAATTTTGAAAACTATPA	1960
Db	89	ATCATTTTATTAATAATTAATTAATTTTGATTTTAACTTATATATATTTATATA	148
QY	1961	ATTTCTAACATATTTATTA - AATATGAGGCTATATATGTAATTTCTTAAT	2015
Db	149	TTTACTTAATTCACATTAATTAATAATTAATTAATTAATAATTAATTAATGAT	208
QY	2020	ATTTTTTTTATTTAGTTATTAATACATTATGACCAATAATATAGTTGGCAATTCAA	2075
Db	209	ACTATTTAGCTATAGTCCAAATTTTAAATTTGATTAATAATATATATATATAT	268
QY	2080	TATCTCATTAATATTTTTCGAATCTCAATATTAATTTAGTCAATAACATGCA	2139
Db	269	TTTCTTAATTAATAATTAATAATACATATATCAATATTAATATTTATTTAATGTTT	328
QY	2140	TAGAAAGTCCAAAAAATTTTGTAAACAGAAACCTCCAAATTTTTTTTTTAAVGA	2195
Db	329	ATTAAA-----ATATATATTTTATTTATTAATAAGATTATTTAATAATATGTA	383
QY	2200	CAAGAATTAACAGATAGAAAACATTTTGTGTGGAATGGAAGTAGTATATACATTAG	2255
Db	384	TTATTTATTTTATTTATATATATATATATTTTAAATATATATATGCTATTTATTTAT	443
QY	2260	CAATTTTAAAAATTTATTAAGCCTATAGCGCTCAAGATGATGTATCTAGTAGGCTA	2319
Db	444	CTTTTATTAAGAAATTTATTAATAATTTTAACTTTAATTTCTTATTTTAAATTTT	503
QY	2320	ATTATATATGCAATGTCGATTCAGAAATGGGACACATGAAGAAACGGAATTAATAT	2379
Db	504	ATATTTATTTATTAATATATATTTCAATTTTATTTATTTATTTATTAATTAATTA	563
QY	2380	AACCTTAAATTAATAATAATTTGACATTAAT	2410
Db	564	TAAATTAATTTATTCATTAATTAATTAATTT	594

RESULT 6
 US-09-641-638-651
 ; Sequence 651, Application US/09641638
 ; Patent No. 6432648
 ; GENERAL INFORMATION:
 ; APPLICANT: Blumenfeld, Marta
 ; APPLICANT: Bougueleret, Lydie
 ; APPLICANT: Chumakov, Ilya
 ; APPLICANT: Cohen, Annick
 ; TITLE OF INVENTION: BIALLELIC MARKERS DERIVED FROM GENOMIC REGIONS CARRYING
 ; TITLE OF INVENTION: GENES INVOLVED IN ARCHIDONIC ACID METABOLISM
 ; FILE REFERENCE: GENSET.05ICPI
 ; CURRENT APPLICATION NUMBER: US/09/641,638
 ; CURRENT FILING DATE: 2000-08-16
 ; PRIOR APPLICATION NUMBER: US 09/502,330
 ; PRIOR FILING DATE: 2000-02-11
 ; PRIOR APPLICATION NUMBER: US 60/133,200
 ; PRIOR FILING DATE: 1999-05-07
 ; PRIOR APPLICATION NUMBER: US 09/275,267
 ; PRIOR FILING DATE: 1999-03-23
 ; PRIOR APPLICATION NUMBER: US 60/119,917

```

? PRIOR FILING DATE: 1999-02-12
? NUMBER OF SEQ ID NOS: 1304
? SOFTWARE: Patent.pm
? SEQ ID NO 651
? LENGTH: 20674
? TYPE: DNA
? ORGANISM: Homo sapiens
? FEATURE:
? NAME/KEY: misc_feature
? LOCATION: 1123..3123
? OTHER INFORMATION: 5'regulatory region
? NAME/KEY: exon
? LOCATION: 3124..3297
? OTHER INFORMATION: exon 1
? NAME/KEY: exon
? LOCATION: 3871..4072
? OTHER INFORMATION: exon 2
? NAME/KEY: exon
? LOCATION: 5552..5633
? OTHER INFORMATION: exon 3
? NAME/KEY: exon
? LOCATION: 5758..5880
? OTHER INFORMATION: exon 4
? NAME/KEY: exon
? LOCATION: 5936..6099
? OTHER INFORMATION: exon 5
? NAME/KEY: exon
? LOCATION: 6349..6509
? OTHER INFORMATION: exon 6
? NAME/KEY: exon
? LOCATION: 7379..7522
? OTHER INFORMATION: exon 7
? NAME/KEY: exon
? LOCATION: 8645..8854
? OTHER INFORMATION: exon 8
? NAME/KEY: exon
? LOCATION: 12254..12340
? OTHER INFORMATION: exon 9
? NAME/KEY: exon
? LOCATION: 12854..13023
? OTHER INFORMATION: exon 10
? NAME/KEY: exon
? LOCATION: 13308..13429
? OTHER INFORMATION: exon 11
? NAME/KEY: exon
? LOCATION: 16567..16667
? OTHER INFORMATION: exon 12
? NAME/KEY: exon
? LOCATION: 16775..16945
? OTHER INFORMATION: exon 13
? NAME/KEY: exon
? LOCATION: 17063..17554
? OTHER INFORMATION: exon 14
? NAME/KEY: misc_feature
? LOCATION: 17555..20674
? OTHER INFORMATION: 3'regulatory region
? NAME/KEY: allele
? LOCATION: 1128
? OTHER INFORMATION: 10-508-191 : polymorphic base C or T
? NAME/KEY: allele
? LOCATION: 1182
? OTHER INFORMATION: 10-508-245 : polymorphic base C or T
? NAME/KEY: allele
? LOCATION: 1559
? OTHER INFORMATION: 10-509-284 : polymorphic base C or T
? NAME/KEY: allele
? LOCATION: 1570
? OTHER INFORMATION: 10-509-295 : deletion of C
? NAME/KEY: allele
? LOCATION: 1827
? OTHER INFORMATION: 10-510-173 : variable motif ATTGA or TTTTTT
? NAME/KEY: allele
? LOCATION: 2048
? OTHER INFORMATION: 10-511-62 : polymorphic base C or T
? NAME/KEY: allele
? LOCATION: 2323
? OTHER INFORMATION: 10-511-337 : insertion of T
? NAME/KEY: allele
? LOCATION: 2341
? OTHER INFORMATION: 10-512-36 : polymorphic base G or C
? NAME/KEY: allele
? LOCATION: 2623
? OTHER INFORMATION: 10-512-318 : polymorphic base A or G
? NAME/KEY: allele
? LOCATION: 2832
? OTHER INFORMATION: 10-513-250 : polymorphic base A or G
? NAME/KEY: allele
? LOCATION: 2844
? OTHER INFORMATION: 10-513-262 : polymorphic base C or T
? NAME/KEY: allele
? LOCATION: 2934
? OTHER INFORMATION: 10-513-352 : polymorphic base A or G
? NAME/KEY: allele
? LOCATION: 2947
? OTHER INFORMATION: 10-513-365 : polymorphic base A or G
? NAME/KEY: allele
? LOCATION: 3802
? OTHER INFORMATION: 12-206-81 : polymorphic base A or G
? NAME/KEY: allele
? LOCATION: 4062
? OTHER INFORMATION: 10-343-231 : deletion of C
? NAME/KEY: allele
? LOCATION: 4088
? OTHER INFORMATION: 12-206-366 : polymorphic base C or T
? NAME/KEY: allele
? LOCATION: 4109
? OTHER INFORMATION: 10-343-278 : polymorphic base C or T
? NAME/KEY: allele
? LOCATION: 4170
? OTHER INFORMATION: 10-343-339 : polymorphic base G or T
? NAME/KEY: allele
? LOCATION: 5903
? OTHER INFORMATION: 10-346-23 : polymorphic base A or G
? NAME/KEY: allele
? LOCATION: 6019
? OTHER INFORMATION: 10-346-141 : polymorphic base A or G
? NAME/KEY: allele
? LOCATION: 6141
? OTHER INFORMATION: 10-346-263 : polymorphic base G or C
? NAME/KEY: allele
? LOCATION: 6183
? OTHER INFORMATION: 10-346-305 : polymorphic base C or T
? NAME/KEY: allele
? LOCATION: 6338
? OTHER INFORMATION: 10-347-74 : polymorphic base A or G
? NAME/KEY: allele
? LOCATION: 6375
? OTHER INFORMATION: 10-347-111 : polymorphic base G or C
? NAME/KEY: allele
? LOCATION: 6429
? OTHER INFORMATION: 10-347-165 : polymorphic base C or T
? NAME/KEY: allele
? LOCATION: 6467
? OTHER INFORMATION: 10-347-203 : polymorphic base A or G
? NAME/KEY: allele
? LOCATION: 6484
? OTHER INFORMATION: 10-347-220 : polymorphic base A or G
? NAME/KEY: allele
? LOCATION: 6534
? OTHER INFORMATION: 10-347-271 : polymorphic base A or T
? NAME/KEY: allele
? LOCATION: 6611
? OTHER INFORMATION: 10-347-348 : polymorphic base A or G
? NAME/KEY: allele
? LOCATION: 7668
? OTHER INFORMATION: 10-348-391 : polymorphic base A or G
```

NAME/KEY: allele
LOCATION: 8608
OTHER INFORMATION: 10-349-47 : polymorphic base C or T
NAME/KEY: allele
LOCATION: 8658
OTHER INFORMATION: 10-349-97 : polymorphic base A or G
NAME/KEY: allele
LOCATION: 8703
OTHER INFORMATION: 10-349-142 : polymorphic base G or C
NAME/KEY: allele
LOCATION: 8777
OTHER INFORMATION: 10-349-216 : deletion of CTG
NAME/KEY: allele
LOCATION: 8785
OTHER INFORMATION: 10-349-224 : polymorphic base G or T
NAME/KEY: allele
LOCATION: 8926
OTHER INFORMATION: 10-349-368 : polymorphic base C or T
NAME/KEY: allele
LOCATION: 12171
OTHER INFORMATION: 10-350-72 : polymorphic base C or T
NAME/KEY: allele
LOCATION: 12429
OTHER INFORMATION: 10-350-332 : polymorphic base C or T
NAME/KEY: allele
LOCATION: 13341
OTHER INFORMATION: 10-507-170 : polymorphic base A or G
NAME/KEY: allele
LOCATION: 13492
OTHER INFORMATION: 10-507-321 : polymorphic base A or C
NAME/KEY: allele
LOCATION: 13524
OTHER INFORMATION: 10-507-353 : polymorphic base C or T
NAME/KEY: allele
LOCATION: 13535

Query Match
Best Local Similarity 52.6%; Score 65.4; DB 4; Length 20674;
Pred. No. 0.0004;
Matches 235; Conservative 0; Mismatches 206; Indels 6; Gaps 4;

QY 1842 ACTGATGATGCTATTTTTCGCAATTTAATAGAAATGCTACTATCAATA 1901
DB 11120 AATATTAATAATTAATTTAATTAATTAATTAATTAATTAATTAATTA 11178
QY 1902 GAAGAAACAGCTGATATTAATTAATTAATTAATTAATTAATTAATTA 1961
DB 11179 ATTAATAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 11238
QY 1962 TTCTACAAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 2021
DB 11239 TTTTAAATTTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 11298
QY 2022 TTTTAAATTTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 2081
DB 11299 TATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 11357
QY 2082 TCTCATTAAT---ATTTTGAATCTGCAATTAATTAATTAATTAATTA 2138
DB 11358 TTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 11417
QY 2139 ATAGAAAGTCCAAAAAAT---TGTAAACAGAAATCTCAAAATTTTAAAT 2197
DB 11418 CTTCATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 11477
QY 2198 AACAAAGAAATTAACAGATAGAAATTTTGTGGAATGGAATGATTAATCA 2257
DB 11478 AAAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 11537
QY 2258 AGCAATTTTAAATTAATTAATTAATTAATTAATTAATTAATTAATTA 2284
DB 11538 AAAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 11564

RESULT 7
US-08-446-855A-1
Sequence 1, Application US/08446855A
Patent No. 5849573
GENERAL INFORMATION:
APPLICANT: Stewart, Thomas S
APPLICANT: Flores, Maria V
APPLICANT: O'Sullivan, William J
TITLE OF INVENTION: Nucleotide sequence encoding cardamyl
TITLE OF INVENTION: phosphate synthetase II
NUMBER OF SEQUENCES: 2
CORRESPONDENCE ADDRESS:
ADDRESSEE: Nixon & Vanderhye PC
STREET: 1100 No. 5849573th Glebe Road, 8th Floor
CITY: Arlington
STATE: Virginia
COUNTRY: USA
ZIP: 22201-4714
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent Release #1.24
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/446,855A
FILING DATE: 06-Jul-1995
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: Mitchell, Leonard C
REGISTRATION NUMBER: 29,009
REFERENCE/DOCKET NUMBER: 47-80
TELECOMMUNICATION INFORMATION:
TELEPHONE: 703-816-4000
TELEFAX: 703-816-4100
INFORMATION FOR SEQ. ID NO. 1:
SEQUENCE CHARACTERISTICS:
LENGTH: 8920 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: genomic
US-08-446-855A-1

Query Match
Best Local Similarity 46.8%; Score 63.6; DB 2; Length 8920;
Pred. No. 0.00074;
Matches 278; Conservative 0; Mismatches 304; Indels 12; Gaps 2;

QY 1874 AATTAGAAATTCATGCTACTATCAATAGAAACAGCTGATTAATTAATTA 1933
DB 152 AAAAATAATAATAATAATAATAATAATAATAATAATAATAATAATAATA 211
QY 1934 AAGACAAATTTTGAATAATGTTAATTAATTAATTAATTAATTAATTA 1992
DB 212 ATATATGATTTATTCATTAATTAATTAATTAATTAATTAATTAATTAAT 271
QY 1993 ATATGATTTTCTATGCTTAATTAATTAATTAATTAATTAATTAATTA 2052
DB 272 TTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 331
QY 2053 GAACCAATATAGTTGGTGAATTCAAATATCTCCATTAATTTTGAATCTCA 2112
DB 332 TATATATATATATATATGTTGGTGTGCTCATTTGTTAATAATTAATTA 391
QY 2113 TATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 2172
DB 392 TATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 451
QY 2173 ACTTCAAAATTTTATTTTATGCA-----ACAGAAATTAACAGATGAAG 2221
DB 452 ATATAGTTTAACTCTTAATTAATAACCAATCCTATTAATTAATTAATTA 511
QY 2222 TATTTTGTGGAATGAGATGATTAATTAATTAATTAATTAATTAATTA 2281

Db 512 TCCCAATATGTGGTCCCTAAATTTATTTATTAATTTATTTATTAATTTATTCATTT 571
 QY 2282 GCCTTACGCGCTCAAAGTATGTATCTAGTAGGGTAATTAATAATGATGGCGATT 234
 Db 572 ATTTATTTTTTTCTTACTATTAATAAAAGTAATTCCTACTAATTTAAAAA 631
 QY 2342 CAGAAATGGGCAACATGAAGGAAATTAATTAATTAATTAATTAATTAATTAAT 240
 Db 632 AAAAAAAAAAAAAAAAAAAAAAAAAATTTACATATGAAGAAATGAACTTGTAATGTA 691
 QY 2402 TGAGTAAATGTCTTCTGACATTTGAGGGCCAAAAAGACAATGCCAAAG 2455
 Db 692 TTTAATAATTTTTTAACATAATATTAATGATATAAAAAAAAAAAAAGAAAAATG 745

```

: RESULT 8
: US-09-150-741-1
: Sequence.1, Application US/09150741
: Patent No. 6183996
: GENERAL INFORMATION:
: APPLICANT: Stewart et al.
: TITLE OF INVENTION: Nucleotide Sequence Encoding Carbamoyl Phosphate
: Patent No. 6183996
: TITLE OF INVENTION: Synthetase II
: FILE REFERENCE:
: CURRENT APPLICATION NUMBER: US/09/150,741
: CURRENT FILING DATE: 1998-09-10
: EARLIER APPLICATION NUMBER: PL6380
: EARLIER FILING DATE: 1992-12-16
: EARLIER APPLICATION NUMBER: AU93/00617
: EARLIER FILING DATE: 1993-12-02
: EARLIER APPLICATION NUMBER: 08/446,855
: EARLIER FILING DATE: 1995-07-06
: NUMBER OF SEQ ID NOS: 15
: SOFTWARE: PatentIn Ver. 2.0
: SEQ ID NO 1
: LENGTH: 8920
: TYPE: DNA
: ORGANISM: Plasmodium falciparum
: US-09-150-741-1

```

	Query Match	2.0%	Score 63.6;	DB 4	Length 8920;	
	Best Local Similarity	46.8%;	Pred. No. 0.0074;			
	Matches 278;	Conservative 0;	Mismatches 304;	Indels 12;	Gaps 2	
OY	1874	AATTGGAATTCAGTCTACTATCATCAATGAAAGAACACGTGAGTATTACATTTTAATTTA	1933			
Db	152	AAAAAAAAATATAAAAAAATAAAAAGATATAAAAAATATTAATTTGATATGATGAT	211			
OY	1934	AAGACAATTTTTGGAAAAATGTTATA-ATTCTACAAATATTTATAAATATGATGCT	1997			
Db	212	ATATAAGATATTATCATTTTATATAACATGATATAAAAAACCTTTTTTTTTTTTTTCT	271			
OY	1993	ATAATGATTTCCATGTTCTTAAAAATTTTTTTTTTATTTAGTGTATAATACATTAT	2052			
Db	272	TTATATTATTAAACATACATTTTAAGTATTTTTATATATATATATATATATATATA	331			
OY	2053	GAACCATAATATAGTTGGTAATTCAAATATCTCCATTATATTTTTTGAAATCTACAAT	2113			
Db	332	TATATATATATATATATGTTTGTGTCTCATTGTGTTATATAAATTAACATAATATAACT	391			
OY	2113	TATTAATTTTAGTACATTAACATACATGATAGAAAGTCCAAAAAATAATTTGTTAACAGA	2172			
Db	392	TATTATATATATTTCCCAATTAATATGAATRCATTTTATATTTTGATGTGACACATTA	451			
OY	2173	ACTTCCAATTTTTTTTTTATGGA-----ACAAGAAATTAACAGATAGAAAC	2221			
Db	452	ATATAGTTTTACCTTCCTTATAAATAAACCATCCATATATTTATACCAATATATAATAC	511			
OY	2222	TATTTGTTGGAATGGAAGTAGAATATACATTAACCAATTTTAAAAATTTATATA	2287			
Db	512	TCCCCAATATTTGCGTCTCTATTAATTTTATTTATATATATTTTATTTATTTATTCATTT	571			

Oy 2282 GCCATACGGCCTCAAGATATGTTATCAGTAGAGGTAAATTAATATATGCAATGCTGGCATT 2341
 Db 572 ATTTATTTTTTTTTCTTACGTTATTAATAAATGTAATTCCTACTAATTTAAAAA
 Oy 2342 CAGAAATGGGACACAATGAAACGGAATTAATAATATTAACCTTAAATAATTAATAAATTT 2401
 Db 632 AAAAAAAAAAAAAAAAAAAAAAAAAAATTTTCATATGAAAAATGACACTTGTTATATGTA 691
 Oy 2402 TGAGTAAATATGTTTTCTTGCATATTTGAGGGGCAAAAAAGACATGCCAAAG 2455
 Db 692 TTTATTAATATTTTAAACATTAATATTAATGTTATTAATAAAAAAAAAAGAAAAATG 745

```

1 RESULT 9
2 US-09-641-638-651/C
3 : Sequence 651: Application US/09641638
4 : Patent No. 6432648
5 : GENERAL INFORMATION:
6 : APPLICANT: Blumenfeld, Marta
7 : APPLICANT: Bouguetieret, Lydie
8 : APPLICANT: Chumakov, Ilya
9 : APPLICANT: Cohen, Annick
10 : TITLE OF INVENTION: BIALLELIC MARKERS DERIVED FROM GENOMIC REGIONS CARRYING
11 : FILE OF INVENTION: GENSET_051CPI
12 : FILE REFERENCE: GENSET_051CPI
13 : CURRENT APPLICATION NUMBER: US/09/641,638
14 : CURRENT FILING DATE: 2000-08-16
15 : PRIOR APPLICATION NUMBER: US 09/502,330
16 : PRIOR FILING DATE: 2000-02-11
17 : PRIOR APPLICATION NUMBER: US 60/133,200
18 : PRIOR FILING DATE: 1999-05-07
19 : PRIOR APPLICATION NUMBER: US 09/275,267
20 : PRIOR FILING DATE: 1999-03-23
21 : PRIOR APPLICATION NUMBER: US 60/119,917
22 : PRIOR FILING DATE: 1999-02-12
23 : NUMBER OF SEQ ID NOS: 1304
24 : SOFTWARE: Patent.pm
25 : SEQ ID NO 651
26 : LENGTH: 20674
27 : TYPE: DNA
28 : ORGANISM: Homo sapiens
29 : FEATURE:
30 : NAME/KEY: misc-feature
31 : LOCATION: 1123..3123
32 : OTHER INFORMATION: 5'regulatory region
33 : NAME/KEY: exon
34 : LOCATION: 3124..3297
35 : OTHER INFORMATION: exon 1
36 : NAME/KEY: exon
37 : LOCATION: 3871..4072
38 : OTHER INFORMATION: exon 2
39 : NAME/KEY: exon
40 : LOCATION: 5552..5633
41 : OTHER INFORMATION: exon 3
42 : NAME/KEY: exon
43 : LOCATION: 5758..5880
44 : OTHER INFORMATION: exon 4
45 : NAME/KEY: exon
46 : LOCATION: 5996..6099
47 : OTHER INFORMATION: exon 5
48 : NAME/KEY: exon
49 : LOCATION: 6349..6509
50 : OTHER INFORMATION: exon 6
51 : NAME/KEY: exon
52 : LOCATION: 7379..7522
53 : OTHER INFORMATION: exon 7
54 : NAME/KEY: exon
55 : LOCATION: 8645..8854
56 : OTHER INFORMATION: exon 8
57 : NAME/KEY: exon
58 : LOCATION: 12254..12340
59 : OTHER INFORMATION: exon 9
60 : NAME/KEY: exon

```


LOCATION: 12854..13023
OTHER INFORMATION: exon 10
NAME/KEY: exon
LOCATION: 13308..13429
OTHER INFORMATION: exon 11
NAME/KEY: exon
LOCATION: 16567..16667
OTHER INFORMATION: exon 12
NAME/KEY: exon
LOCATION: 16775..16945
OTHER INFORMATION: exon 13
NAME/KEY: exon
LOCATION: 17063..17554
OTHER INFORMATION: exon 14
NAME/KEY: misc-feature
LOCATION: 17555..20674
OTHER INFORMATION: 3' regulatory region
NAME/KEY: allele
LOCATION: 1128
OTHER INFORMATION: 10-508-191 : polymorphic base C or T
NAME/KEY: allele
LOCATION: 1182
OTHER INFORMATION: 10-508-245 : polymorphic base C or T
NAME/KEY: allele
LOCATION: 1559
OTHER INFORMATION: 10-509-284 : polymorphic base C or T
NAME/KEY: allele
LOCATION: 1570
OTHER INFORMATION: 10-509-295 : deletion of C
NAME/KEY: allele
LOCATION: 1827
OTHER INFORMATION: 10-510-173 : variable motif ATTGA or TTTTTT
NAME/KEY: allele
LOCATION: 2048
OTHER INFORMATION: 10-511-62 : polymorphic base C or T
NAME/KEY: allele
LOCATION: 2323
OTHER INFORMATION: 10-511-337 : insertion of T
NAME/KEY: allele
LOCATION: 2341
OTHER INFORMATION: 10-512-36 : polymorphic base G or C
NAME/KEY: allele
LOCATION: 2623
OTHER INFORMATION: 10-512-318 : polymorphic base A or G
NAME/KEY: allele
LOCATION: 2832
OTHER INFORMATION: 10-513-250 : polymorphic base A or G
NAME/KEY: allele
LOCATION: 2844
OTHER INFORMATION: 10-513-262 : polymorphic base C or T
NAME/KEY: allele
LOCATION: 2934
OTHER INFORMATION: 10-513-352 : polymorphic base A or G
NAME/KEY: allele
LOCATION: 2947
OTHER INFORMATION: 10-513-365 : polymorphic base A or G
NAME/KEY: allele
LOCATION: 3802
OTHER INFORMATION: 12-206-81 : polymorphic base A or G
NAME/KEY: allele
LOCATION: 4062
OTHER INFORMATION: 10-343-231 : deletion of C
NAME/KEY: allele
LOCATION: 4088
OTHER INFORMATION: 12-206-366 : polymorphic base C or T
NAME/KEY: allele
LOCATION: 4109
OTHER INFORMATION: 10-343-278 : polymorphic base C or T
NAME/KEY: allele
LOCATION: 4170
OTHER INFORMATION: 10-343-339 : polymorphic base G or T
NAME/KEY: allele
LOCATION: 5903

OTHER INFORMATION: 10-346-23 : polymorphic base A or G
NAME/KEY: allele
LOCATION: 6019
OTHER INFORMATION: 10-346-141 : polymorphic base A or G
NAME/KEY: allele
LOCATION: 6141
OTHER INFORMATION: 10-346-263 : polymorphic base G or C
NAME/KEY: allele
LOCATION: 6183
OTHER INFORMATION: 10-346-305 : polymorphic base C or T
NAME/KEY: allele
LOCATION: 6338
OTHER INFORMATION: 10-347-74 : polymorphic base A or G
NAME/KEY: allele
LOCATION: 6375
OTHER INFORMATION: 10-347-111 : polymorphic base G or C
NAME/KEY: allele
LOCATION: 6429
OTHER INFORMATION: 10-347-165 : polymorphic base C or T
NAME/KEY: allele
LOCATION: 6467
OTHER INFORMATION: 10-347-203 : polymorphic base A or G
NAME/KEY: allele
LOCATION: 6484
OTHER INFORMATION: 10-347-220 : polymorphic base A or G
NAME/KEY: allele
LOCATION: 6534
OTHER INFORMATION: 10-347-271 : polymorphic base A or T
NAME/KEY: allele
LOCATION: 6611
OTHER INFORMATION: 10-347-348 : polymorphic base A or G
NAME/KEY: allele
LOCATION: 7668
OTHER INFORMATION: 10-348-391 : polymorphic base A or G
NAME/KEY: allele
LOCATION: 8608
OTHER INFORMATION: 10-349-47 : polymorphic base C or T
NAME/KEY: allele
LOCATION: 8658
OTHER INFORMATION: 10-349-97 : polymorphic base A or G
NAME/KEY: allele
LOCATION: 8703
OTHER INFORMATION: 10-349-142 : polymorphic base G or C
NAME/KEY: allele
LOCATION: 8777
OTHER INFORMATION: 10-349-216 : deletion of CTG
NAME/KEY: allele
LOCATION: 8785
OTHER INFORMATION: 10-349-224 : polymorphic base G or T
NAME/KEY: allele
LOCATION: 8926
OTHER INFORMATION: 10-349-368 : polymorphic base C or T
NAME/KEY: allele
LOCATION: 12171
OTHER INFORMATION: 10-350-72 : polymorphic base C or T
NAME/KEY: allele
LOCATION: 12429
OTHER INFORMATION: 10-350-332 : polymorphic base C or T
NAME/KEY: allele
LOCATION: 13341
OTHER INFORMATION: 10-507-170 : polymorphic base A or G
NAME/KEY: allele
LOCATION: 13492
OTHER INFORMATION: 10-507-321 : polymorphic base A or C
NAME/KEY: allele
LOCATION: 13524
OTHER INFORMATION: 10-507-353 : polymorphic base C or T
NAME/KEY: allele
LOCATION: 13535

Query Match 2.0%; Score 63; DB 4; Length 20674;
Best Local Similarity 48.0%; Pred. No. 0.0012;
Matches 224; Conservative 0; Mismatches 230; Indels 13; Gaps 1;

QY 1825 TAATGACAGCCTCTACACAGTGTGATGCAATTTTCTGCAATATTAATTTGCAATT 1884
DB 11554 TAAATATTAAATATTTCTTACGCTAATTAATTAATTAATTAATTAATTAATTAAT 11495
QY 1885 CAATGCTACTACATCAATGAGAAACAGCTGATTTTCAATTTTAAATTAAGACAAAT 1944
DB 11494 ATTAATATTTAATTTAATTAATTAATTTAATTAATTAATTAATTAATTAATTAAT 11435
QY 1945 TTGAAAAAATGTTAATTTCTTACAAATATTAATTAATTAATTAATTAATTAATTTTC 2004
DB 11434 TTAATTTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 11375
QY 2005 CTATGTTCTTAAATATTTTCTTATTAATTTAGTTAATTAATTAATTAATTAATTAAT 2064
DB 11374 AATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 11315
QY 2065 GTTGCAATTCATCAATATC-----TCCATTAATTTTGTGAATCTACAA 2111
DB 11314 TTAAGTTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 11255
QY 2112 TTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 2171
DB 11254 ATATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 11195
QY 2172 AACTCCAAATTTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCT 2231
DB 11194 TTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 11135
QY 2232 TCGAATGCAAGTAGTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 2278
DB 11134 TTTAATTTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 11088

RESULT 10
US-08-817-926-27
Sequence 27, Application US/08817926
Patent No. 6001590
GENERAL INFORMATION:
APPLICANT: Komeda, Toshinori
APPLICANT: Suda, Hisako
APPLICANT: Tamai, Yukio
APPLICANT: Imatsu, Akihito
APPLICANT: Kato, No. 6001590uo
APPLICANT: Sakai, Yasuyoshi
TITLE OF INVENTION: PROMOTER/TERMINATOR FOR CANDIDA BOIDINI
TITLE OF INVENTION: FORMATE DEHYDROGENASE GENE
NUMBER OF SEQUENCES: 51
CORRESPONDENCE ADDRESS:
ADDRESSEE: Foley & Lardner
STREET: 3000 K Street, N.W., Suite 500
CITY: Washington
STATE: D.C.
COUNTRY: USA
ZIP: 20007-5109
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentlin Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/817,926
FILING DATE: 09-MAY-1997
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: PCT/JP96/02597
FILING DATE: 12-SEP-1996
PRIOR APPLICATION DATA: JP 234133/1995
FILING DATE: 12-SEP-1995
APPLICATION NUMBER: JP 42536/1996
FILING DATE: 29-FEB-1996

ATTORNEY/AGENT INFORMATION:
NAME: Bent, Stephen A.
REGISTRATION NUMBER: 29,768
REFERENCE/DOCKET NUMBER: 081356/0112
TELEPHONE: (202)672-5300
TELEFAX: (202)672-5399
TELEX: 904136
INFORMATION FOR SEQ ID NO: 27:
SEQUENCE CHARACTERISTICS:
LENGTH: 4818 base pairs
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: linear
MOLECULE TYPE: DNA (genomic)
ORIGINAL SOURCE:
ORGANISM: Candida boidini
US-08-817-926-27

Query Match 2.08; Score 62.6; DB 3; Length 4818;
Best Local Similarity 44.4%; Pred. No. 0.001;
Matches 338; Conservative 0; Mismatches 419; Indels 4; Gaps 2;

QY 1525 ACTTTAATTTGCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCT 1584
DB 4020 AATCTTTAATTTGATTTCTTTTATTTGATTTCTTCAATTTCTTGATTAATTAATTTGTTG 4079
QY 1585 AAGAGCGAATTTGTAACATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 1644
DB 4080 TTGATTTCTTCAATTTGATTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTT 4138
QY 1645 TACATCTTTGATTTGATTTGATTTGATTTGATTTGATTTGATTTGATTTGATTTGATTT 1704
DB 4139 TATTTCTTTGTTTCTTCAATTTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 4198
QY 1705 CGAGGTAAATTTGCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCT 1764
DB 4199 ATTTATTTATTTATTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTT 4258
QY 1765 AATTTTGAATTTGTAACCTTTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCT 1821
DB 4259 CTGAGTTAATTTGATTTGATTTGATTTGATTTGATTTGATTTGATTTGATTTGATTTGATTT 4318
QY 1822 CCATTAATGACCTCTTCAACCTGATTTGATTTGATTTGATTTGATTTGATTTGATTTGATTT 1881
DB 4319 TTAATTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTT 4378
QY 1882 ATTCATGCTACTATCAATGAGAAACAGCTGATTTGATTTGATTTGATTTGATTTGATTTGATTT 1941
DB 4379 CTTTGTTGTTATTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTT 4438
QY 1942 ATTTTGAATTTGATTTGATTTGATTTGATTTGATTTGATTTGATTTGATTTGATTTGATTT 2001
DB 4439 ATTTTATTTATTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTT 4498
QY 2002 TTCCATGTTCTTAAATTTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTT 2061
DB 4499 TTTCTAATTTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTT 4558
QY 2062 ATGAGTGGTGAATTCATTAATTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTT 2121
DB 4559 TTTGTTAATTTGTTATTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTT 4618
QY 2122 TTAATCAATTTGATTTGATTTGATTTGATTTGATTTGATTTGATTTGATTTGATTTGATTTGATTT 2181
DB 4619 TTAATTTGAGAAATTTGATTTGATTTGATTTGATTTGATTTGATTTGATTTGATTTGATTTGATTT 4678
QY 2182 TTTTCTT 2241
DB 4679 TGACATATTTGAGAAATTTGATTTGATTTGATTTGATTTGATTTGATTTGATTTGATTTGATTTGATTT 4738
QY 2242 GTAGTATTTACTTTGAGCAATTTTAAATTTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 2282

Db 4739 TAACTAAGCTAAAGCTAAATTTCTATTTTAAAG 4779

RESULT 11

US-07-867-106-2/c

Sequence 2, Application US/07867106

Patent No. 5389526

GENERAL INFORMATION:

APPLICANT: Slade, Martin B

APPLICANT: Chang, Andy C M

APPLICANT: Williams, Keith L

TITLE OF INVENTION: Improved Plasmid Vectors for Cellular

NUMBER OF INVENTION: Slime Moulds of the Genus Dictyostelium

NUMBER OF SEQUENCES: 19

CORRESPONDENCE ADDRESSES:

ADDRESSEE: Woodcock Washburn Kurtz Mackiewicz & No. 5389526r1s

STREET: One Liberty Place 46th Floor

CITY: Philadelphia

STATE: PA

COUNTRY: USA

ZIP: 19103

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk

OPERATING SYSTEM: IBM PC compatible

SOFTWARE: Patent Release #1.0, Version #1.25

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/07/867,106

FILING DATE: 19920625

PRIOR APPLICATION DATA:

APPLICATION NUMBER: AU PU 7187

APPLICATION NUMBER: PCT/AU90/00530

FILING DATE: 02-NOV-1989

ATTORNEY/AGENT INFORMATION:

NAME: Feeney, Joanne Longo

REGISTRATION NUMBER: 35,134

REFERENCE/DOCKET NUMBER: RICE-0002

TELEPHONE: 215-568-3100

TELEFAX: 215-568-3439

INFORMATION FOR SEQ ID NO: 2:

SEQUENCE CHARACTERISTICS:

LENGTH: 5852 base pairs

TYPE: NUCLEIC ACID

STRANDEDNESS: single

TOPOLOGY: linear

MOLECULE TYPE: DNA (genomic)

ANTI-SENSE: NO

FEATURE:

NAME/KEY: CDS

LOCATION: 2378..5038

FEATURE:

NAME/KEY: CDS

LOCATION: 2378..5038

NAME/KEY: CDS

LOCATION: 2378..5038

US-07-867-106-2

Query Match

Best Local Similarity 1.9%; Score 61.8; DB 1; Length 5852;

Matches 209; Conservative 0; Mismatches 217; Indels 7; Gaps 1;

Db 1767 TTTTGAAGATGTACCTTTTACTCATAGTATTAACCGTAGTTTGGTGCATA 1826

Db 5787 TTTAATGTTATTTATTTATTTATTTGTTGTTATTTTATATATATGTTAT 5728

Db 1827 ATGACAGCCTTACACTGTGATAGTCAATTTTCTGCAATATTAATTAAGCAATTC 1886

Db 5727 TGTGTTGTTTATTTTACTTATTTTCAATTTTATTAATTAATTAATTAATTT 5668

Db 1887 ATGCTACATCATTAAGAAAGACAGCTAGTATTAATTAATTAAGCAATTTT 1946

Db 5667 TAAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 5608

Db 1947 TGAATAATGTTATTAATTTCTAACAATATTAATTAATTAATTAATTAATTAAT 2006

Db 5607 AAAAAAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 5548

Db 2007 ATGCTCTAATAATTTTATTTTATTA-----TTAGTTATTAATTAATTAATTA 2059

Db 5547 AAAAAATTAATTAATTTTATTTTATTTTATTTTATTTTATTTTATTTTATTT 5488

Db 2060 TAAATGTTGCTGATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 2119

Db 5487 TAAACATATTTGATTTTATTTTATTTTATTTTATTTTATTTTATTTTATTTT 5428

Db 2120 ATTTACTCAATTAACATGATAGTAAGTTCCAAAAATTTTGTATACAGAACTTCA 2179

Db 5427 TAAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 5368

Db 2180 AATTTTATTTTATTTTATTTTATTTTATTTTATTTTATTTTATTTTATTT 2192

Db 5367 AATTTGATGATCT 5355

RESULT 12

US-09-316-083-2

Sequence 2, Application US/09316083A

Patent No. 6280942

GENERAL INFORMATION:

APPLICANT: The Institute of Physical and Chemical Research

TITLE OF INVENTION: Endonuclease

FILE REFERENCE: PH-651

CURRENT APPLICATION NUMBER: US/09/316,083A

CURRENT FILING DATE: 1999-05-20

EARLIER APPLICATION NUMBER: JP98/141861

NUMBER OF SEQ ID NOS: 38

SOFTWARE: Patent Ver. 2.0

SEQ ID NO 2

LENGTH: 1431

TYPE: DNA

ORGANISM: Saccharomyces cerevisiae

FEATURE:

NAME/KEY: CDS

LOCATION: (1)..(1428)

US-09-316-083-2

Query Match

Best Local Similarity 1.9%; Score 61.2; DB 4; Length 1431;

Matches 215; Conservative 0; Mismatches 238; Indels 1; Gaps 1;

Db 1899 ATAGAGAAGACAGCTGATTTATTAATTTAAGACAAATTTTGAATAATGTTA 1958

Db 164 ATAAATTAATTAATTTATTTATTTATTTATTTATTTATTTATTTATTTATTT 223

Db 1959 TAAATTTCTAACAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 2018

Db 224 AACACGTAACACCTGATGCTAATTAATTAATTAATTAATTAATTAATTAAT 283

Db 2019 TATTTTATTTATTTATTTATTTATTTATTTATTTATTTATTTATTTATTT 2078

Db 284 AATTTATTAATTAATTTATTTATTTATTTATTTATTTATTTATTTATTTATTT 343

Db 2079 ATATCTCCATTAATTAATTTTGAATCTACA-AATTTAATTAATTTTGAATCAATG 2137

Db 344 ATAAATTAATTAATTAATTTATTTATTTATTTATTTATTTATTTATTTATTT 403

Db 2138 CATGAAGAGTTCCAAAAATTTTGTAAACAGAACTTCCAAATTTTATTTATG 2197

Db 404 ATTTAATTTAATTAATTAATTTAATTTAATTTAATTTAATTTAATTTAATTT 463

Db 2198 AACAGAAATTAACGATAGAAACTATTTTGTGGAAGAGTAATTAATTAATTA 2257

Db 464 ATAAATTTATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 523

Db 2258 AGCAATTTTAAAAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 2317


```

?      REGISTRATION NUMBER: 38,241
?      REFERENCE/DOCKET NUMBER: PF/5-30306/A/CGL1976
?      TELECOMMUNICATION INFORMATION:
?      TELEPHONE: 919-541-8587
?      TELEFAX: 919-541-8689
?      INFORMATION FOR SEQ ID NO: 288:
?      SEQUENCE CHARACTERISTICS:
?      LENGTH: 837 base pairs
?      TYPE: nucleic acid
?      STRANDEDNESS: single
?      TOPOLOGY: linear
?      MOLECULE TYPE: DNA (genomic)
?      ORIGINAL SOURCE:
?      ORGANISM: PAGI241RP
?      GS-08-998-416-288

```

Query Match	1.9%	Score 60.6;	DB 4;	Length 837;
Best Local Similarity	49.1%	Pred. NC. 0.0017;		
Matches 191; Conservative	0;	Mismatches 194;	Indels 4;	Gaps 1;

QY	1939	AAATTTTTGGAAAATGTTATAATTTCTCAACAAATATTATTAANAATGATCGCTTAAAG	1998
Dd	608	ATATATTTTATTTAATTAATTAATAATGATGAATAATATTATTAATAATTAATTTAAAT	549
QY	1999	TATTCCTANGTCTTAAAAATATTTTTTTTTTATATTAGTTATTAATACATTATGAMCCA	2056
Dd	548	AAAAATTAATAATTAATGAATATTAATTTTATTAATTAATTAATAAATAATTAATAATA	489
QY	2059	ATATACTTGGTGAATTCAAATATCCTCATTAATTTTTTGGAAATGCACAATATTAA	2116
Dd	488	GAAATTTAAAGTTAAAAATTAATTTAATTAATTAATCTTTAAAAAGATTAAATTAATTA	429
QY	2119	TATTTAGCATATCAATGCAATGATAGAAAGTCCAAAAAANNTTGTATACAGAATCTCC	2176
Dd	428	TCACATATATRTTTATTAATAATTAACATATTATTAATTAATAAATAATTAATTAACATATT	369
QY	2179	AAATTTTTTTTTTTATGGAACAAGAAATAACAGATAGAAAACATATTTTGGTGTGAAATG	2238
Dd	368	TAAATTTAACTTTATATATATTAATAAATATATATTTTATTAACACATTAAT---AAATA	313
QY	2239	GAACTAGTAAATATCAATTAGCAAATTTTAAAAAATTTATTAAGCTATACGCCGTCAA	2296
Dd	312	ATATATTAATTAATGATATATCTATTTTAATTAATTTATTAAGAAAATATTAATATCTAATA	253
QY	2299	GTATGTATCTAGTAGGTATTAATA	2327
Dd	252	ATATTTTAATTAACATTTTAAAAATTGAA	224

RESULT 15
 US-08-446-855A-1/c
 : Sequence 1, Application US/08446855A
 : Patent No. 5849573
 :
 : GENERAL INFORMATION:
 :
 : APPLICANT: Stewart, Thomas S
 : APPLICANT: Flores, Maria V
 : APPLICANT: O'Sullivan, William J
 : TITLE OF INVENTION: Nucleotide sequence encoding c
 : TITLE OF INVENTION: phosphatase synthetase II
 :
 : NUMBER OF SEQUENCES: 2
 :
 : CORRESPONDENCE ADDRESS:
 : ADDRESS: Nixon & Vanderhye PC
 : STREET: 1100 No. 5849573th Glabe Road, 8th Floor
 : CITY: Arlington
 : STATE: Virginia
 :
 : COUNTRY: USA
 :
 : ZIP: 22201-4714
 :
 : COMPUTER READABLE FORM:
 : MEDIUM TYPE: Floppy disk
 : COMPUTER: IBM PC compatible.
 : OPERATING SYSTEM: PC-DOS/MS-DOS
 : SOFTWARE: Patentln Release #1.24
 :
 : CURRENT APPLICATION DATA:

```

1      APPLICATION NUMBER:  US/08/446,855A
2      FILING DATE:  06-Jul-1995
3      CLASSIFICATION:  435
4      ATTORNEY/AGENT INFORMATION:
5      NAME:  Miltchard, Leonard C
6      REGISTRATION NUMBER:  29,009
7      REFERENCE/DOCKET NUMBER:  47-80
8      TELECOMMUNICATION INFORMATION:
9      TELEPHONE:  703-816-4000
10     TELEFAX:  703-816-4100
11     INFORMATION FOR SEQ ID NO:  1:
12     SEQUENCE CHARACTERISTICS:
13     LENGTH:  8920 base pairs
14     TYPE:  nucleic acid
15     STRANDEDNESS:  single
16     TOPOLOGY:  linear
17     MOLECULE TYPE:  genomic
18     US-08-446-855A-1

```

Query Match	1.9%	Score	60.4	DB	2	Length	8920
Best Local Similarity	48.8%	Pred. No.	0.0031				
Matches	163	Conservative	0	Mismatches	171	Indels	0
						Gaps	0

QY	1919	TTACATATTTTAAATTAAGACAAAAATTTTGGAAAATGTTATTAATTTCCAAACAATTTTAT	1978
Db	8751	TTTTATTTTTTTTAAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT	8692
QY	1979	AAAAATGAGCCATPAATGATATTTCCATGTCTCTTAAATATTTTTTTTTTATATTTAGT	2038
Db	8691	AAACCATTTGGTTATATACATATGAGAAATTAATTAATTTGTTATATTAACAAAATTT	8633
QY	2039	TATATATACATTTAGAACCAATTAATAGTGGTGAATTCGAATATCTCATTAATTAATTTAT	2098
Db	8631	TATATAAATAAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT	8572
QY	2099	TGAAATCTACAAATTTATTAATATTAGTCATTAACATATGCATGAGAAAGTTCGAAAAAAA	2155
Db	8571	TTAAACGTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT	8511
QY	2159	TTTTTGTAAACGAACCTCCAAATTTTTTTTTTTTTTAAACAGCAAGAAATTAACAGATAGAA	2218
Db	8511	TATATATATATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT	8455
QY	2219	AACATATTTTGTGTGGAATGGAAGTAGTAATATA	2252
Db	8451	AAATACATAAATAAGTAATGTAAATCGCAATTTA	8418

Search completed: June 11, 2003, 01:53:50
Job time : 155.509 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: June 11, 2003, 01:50:19 ; Search time 459.588 Seconds
(without alignments)
9750.738 Million cell updates/sec

Title: US-09-502-426b-1_COPY_1_3202

Perfect score: 3202
Sequence: 1 atcttggtatattatgtt.....gagagagagaactagctcc 3202

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 870385 segs, 699768693 residues

Total number of hits satisfying chosen parameters: 1740770

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : Published Applications NA:*

1: /cgn2_6/ptodata/1/pubpna/US07_PUBCOMB.seq:*
2: /cgn2_6/ptodata/1/pubpna/PCCT_NEW_PUB.seq:*
3: /cgn2_6/ptodata/1/pubpna/US06_NEW_PUB.seq:*
4: /cgn2_6/ptodata/1/pubpna/US06_PUBCOMB.seq:*
5: /cgn2_6/ptodata/1/pubpna/US07_NEW_PUB.seq:*
6: /cgn2_6/ptodata/1/pubpna/PCBUS_PUBCOMB.seq:*
7: /cgn2_6/ptodata/1/pubpna/US08_NEW_PUB.seq:*
8: /cgn2_6/ptodata/1/pubpna/US08_PUBCOMB.seq:*
9: /cgn2_6/ptodata/1/pubpna/US09_NEW_PUB.seq:*
10: /cgn2_6/ptodata/1/pubpna/US09_PUBCOMB.seq:*
11: /cgn2_6/ptodata/1/pubpna/US10_NEW_PUB.seq:*
12: /cgn2_6/ptodata/1/pubpna/US10_PUBCOMB.seq:*
13: /cgn2_6/ptodata/1/pubpna/US60_NEW_PUB.seq:*
14: /cgn2_6/ptodata/1/pubpna/US60_PUBCOMB.seq:*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	91.6	2.9	960	US-10-198-846-6381	Sequence 6381, Ap
2	82.6	2.6	15732	US-10-239-676-95	Sequence 95, Appl
3	81.6	2.5	5689	US-10-239-676-90	Sequence 90, Appl
4	80	2.5	640681	US-09-790-988-1	Sequence 1, Appl
5	79.4	2.5	960	US-10-198-846-6381	Sequence 6381, Ap
6	75.2	2.3	5979	US-10-239-676-18	Sequence 18, Appl
7	73.4	2.3	5689	US-10-239-676-90	Sequence 90, Appl
8	73	2.3	5979	US-10-239-676-18	Sequence 18, Appl
9	71.2	2.2	15732	US-10-239-676-95	Sequence 95, Appl
10	71	2.2	12405	US-10-239-676-35	Sequence 35, Appl
11	68.8	2.1	335913	US-09-754-853A-2	Sequence 2, Appl
12	68.8	2.1	335913	US-09-754-853A-3	Sequence 3, Appl
13	68	2.1	6298	US-10-239-676-64	Sequence 64, Appl
14	68	2.1	11260	US-10-239-676-20	Sequence 20, Appl
15	67.4	2.1	5917	US-10-087-464-9	Sequence 9, Appl
16	67.2	2.1	4985	US-10-094-240-10	Sequence 10, Appl
17	67.2	2.1	8588	US-10-239-676-178	Sequence 178, Appl
18	67	2.1	4985	US-10-094-240-10	Sequence 10, Appl
19	66.6	2.1	640681	US-09-790-988-1	Sequence 1, Appl

20	66.4	2.1	424	10	US-09-960-352-11218	Sequence 11218, A
21	66.4	2.1	7089	9	US-10-239-676-67	Sequence 67, Appl
22	66.4	2.1	7906	9	US-10-239-676-97	Sequence 97, Appl
23	65.8	2.1	3991	9	US-10-074-045-60	Sequence 60, Appl
24	65.8	2.1	7195	9	US-10-239-676-30	Sequence 30, Appl
25	65.2	2.0	11812	9	US-10-239-676-210	Sequence 210, Appl
26	65.2	2.0	335913	9	US-09-754-853A-2	Sequence 2, Appl
27	65.2	2.0	335913	9	US-09-754-853A-3	Sequence 3, Appl
28	64.8	2.0	6844	9	US-10-239-676-46	Sequence 46, Appl
29	64.8	2.0	6620	9	US-10-239-676-195	Sequence 195, Appl
30	64.4	2.0	473	10	US-09-960-352-11212	Sequence 11212, A
31	64.2	2.0	419	10	US-10-239-676-206	Sequence 206, Appl
32	64.2	2.0	17142	9	US-10-239-676-159	Sequence 159, Appl
33	64	2.0	9515	9	US-10-239-676-28	Sequence 28, Appl
34	63.8	2.0	17848	9	US-10-239-676-102	Sequence 102, Appl
35	63.6	2.0	11836	9	US-10-239-676-102	Sequence 102, Appl
36	63.2	2.0	6620	9	US-10-239-676-196	Sequence 196, Appl
37	63.2	2.0	11836	9	US-10-239-676-102	Sequence 102, Appl
38	63	2.0	11812	9	US-10-239-676-210	Sequence 210, Appl
39	62.6	2.0	12968	9	US-10-239-676-202	Sequence 202, Appl
40	62.6	2.0	15732	9	US-10-239-676-95	Sequence 95, Appl
41	62.4	1.9	6158	9	US-10-239-676-24	Sequence 24, Appl
42	62.2	1.9	6203	9	US-10-239-676-184	Sequence 184, Appl
43	62.2	1.9	6298	9	US-10-239-676-64	Sequence 64, Appl
44	62.2	1.9	11812	9	US-10-239-676-209	Sequence 209, Appl
45	62.2	1.9	127197	9	US-09-754-853A-1	Sequence 1, Appl

ALIGNMENTS

RESULT 1
US-10-198-846-6381/C
Sequence 6381, Application US/10198846
Publication No. US2003009974A1
GENERAL INFORMATION:
APPLICANT: Lillie, James
APPLICANT: Xu, Yongyao
APPLICANT: Wang, Youzhen
APPLICANT: Steinhann, Kathleen
TITLE OF INVENTION: NOVEL GENES, COMPOSITIONS, KITS, AND METHODS
TITLE OF INVENTION: FOR IDENTIFICATION, ASSESSMENT, PREVENTION, AND
FILE REFERENCE: THERAPY OF BREAST CANCER
CURRENT FILING DATE: 2002-07-18
CURRENT APPLICATION NUMBER: US/10/198, 846
PRIOR APPLICATION NUMBER: 60/306, 220
PRIOR FILING DATE: 2001-07-18
NUMBER OF SEQ ID NOS: 14084
SOFTWARE: FastSeq for Windows Version 4.0
SEQ ID NO 6381
LENGTH: 960
TYPE: DNA
ORGANISM: Homo sapiens
FEATURE:
NAME/KEY: misc.feature
LOCATION: 1, 8, 66, 71, 79, 81, 84, 85, 86, 88, 117, 119, 125, 134,
LOCATION: 149, 156, 162, 166, 172, 179, 180, 194, 230, 249, 250, 253,
LOCATION: 268, 271, 273, 274, 275, 276, 277, 286, 291, 299, 312, 333,
LOCATION: 359, 366, 374, 393, 404, 406, 409, 411, 422, 424, 427
OTHER INFORMATION: n = A,T,C or G
FEATURE:
NAME/KEY: misc.feature
LOCATION: 428, 433, 450, 465, 468, 471, 487, 510, 512, 513, 516,
LOCATION: 517, 518, 528, 530, 534, 539, 553, 555, 559, 575, 580, 588,
LOCATION: 590, 610, 625, 626, 642, 645, 656, 658, 659, 675, 685, 686,
LOCATION: 690, 700, 702, 706, 716, 731, 745, 748, 756, 798, 802
OTHER INFORMATION: n = A,T,C or G
FEATURE:
NAME/KEY: misc.feature
LOCATION: 807, 808, 819, 820, 821, 830, 838, 840, 844, 845, 849, 861,
LOCATION: 871, 874, 875, 891, 882, 895, 898, 900, 901, 902, 904, 907,
LOCATION: 917, 923, 925, 926, 927, 929, 935

QY	1996	ATGATATTCCTAATGCTTAAATATTTTATTTATTTAGTATTAATAACATTAAGAA	2055
Db	6234	ATTATATTTATAAATATATATCTATTTATTACTTTATTAATAATACATATTAANAATTTTTC	6175
QY	2056	CCAATAATAGTTGGTGAATTCAAATATCTCCATTAATATTTTGAATCTACAAATTAAT	2115
Db	6174	CTCTCTTAACACACACATTAACATTAATTAAGTAAATTAACATTAATTAATTAATTAAT	6115
QY	2116	TAATATTTAGTCATATACATGCTATAGAAGTTCACAAAAAATTTGGTATACGAAC	2175
Db	6114	AATTAATAATCAAAAAAACTCAATTAATAAATAAACAACATTAATTAATTAATAAT	6055
QY	2176	TCCAAATTTTTTTTTTTATGAGAACAGAAATAACAGATAGAAACATTTTGGTGGGA	2235
Db	6054	TCACACATCTCTCAATTACTTAATAAAAAACAAAAATAAATAAATTAATTTAAATTTAA	5995
QY	2236	ATGGAAGTAGTAATATACATTAAGCAAAATTTTAAAAAATTAATTAAGCTATACGGCTC	2295
Db	5994	ATTTTTCAAACATTAATTAACAAACTTCTAAAAAACTTAATCTTAATTAATTAATTT	5935
QY	2296	AAAGTATGTTATCTAGTAGGTGTAATTAATATGCAATGAGCGGCTTCAGATTTGGGCAA	2355
Db	5934	AAAATCCACGTTTATTAATCTCTCTCAAAATACTCAATAAATTAATTAATAAAACCTTATA	5875
QY	2356	CAATGAAGAGGAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT	2415
Db	5874	TATCAAAACCATTAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATCT	5815
QY	2416	TTCTGACTATTGA 2428	
Db	5814	CCGATCTTTTCA 5802	

RESULT 3
 US-10-239-676-90/c
 Sequence 90, Application US/10239676
 Publication No. US20030082609a1
 GENERAL INFORMATION:
 APPLICANT: OLEK, Alexander
 APPLICANT: PIEPENBROCK, Christian
 APPLICANT: BERLIN, Kurt
 TITLE OF INVENTION: Diagnosis of Diseases Associated with Gene Regulation
 FILE REFERENCE: 5013.1003
 CURRENT APPLICATION NUMBER: US/10/239, 676
 CURRENT FILING DATE: 2002-09-24
 PRIOR APPLICATION NUMBER: PCT/EP01/03968
 DE 10019058.8
 DE 10019173.8
 DE 10032529.7
 DE 10043826.1
 PRIOR FILING DATE: 2001-04-06
 2000-04-06
 2000-04-07
 2000-06-30
 2000-09-01
 NUMBER OF SEQ ID NOS: 228
 SEQ ID NO 90
 LENGTH: 5689
 TYPE: DNA
 ORGANISM: Artificial Sequence
 FEATURE:
 OTHER INFORMATION: chemically treated genomic DNA (Homo sapiens)
 NAME/KEY: unsure
 LOCATION: (1818, 1930..1931, 1934..1935, 1947..1948, 1966)
 US-10-239-676-90

Query Match	2.5%	Score 81.6;	DB 9;	Length 5689;
Best Local Similarity	49.1%;	Pred. No. 0.00049;		
Matches 285; Conservative.	0;	Mismatches 292;	Indels 3;	Gaps 3;

1824 ATATGACAGCCTTACAACGTGTGATGCATTTTTCCTGCCAATATTTAAATTAATGGCAT 1883

[illegible]

```

RESULT 4
US-09-790-988-1
Sequence 1, Application US/09790988
Patent No. US20020127687A1
GENERAL INFORMATION:
APPLICANT: SHIGENOBU, SHUJI
APPLICANT: MATANABE, HIDEMI
APPLICANT: HATTORI, MASAHIRA
APPLICANT: SAKAKI, YOSHIYUKI
TITLE OF INVENTION: GENOME DNA OF BACTERIAL SYMBIONT OF APHIDS
FILE REFERENCE: 081356/0159
CURRENT APPLICATION NUMBER: US/09/790,988
CURRENT FILING DATE: 2001-02-23
PRIOR APPLICATION NUMBER: JP2000-107160
PRIOR FILING DATE: 2000-04-07
NUMBER OF SEQ ID NOS: 7
SOFTWARE: PatentIn Ver. 2.1
SEQ ID NO 1
LENGTH: 640681
TYPE: DNA
ORGANISM: Buchnera sp.
US-09-790-988-1

```

	Query Match	Similarity	2.5%	Score	80:	DB	10:	Length	640681:
	Best Local	Similarity	45.8%	Pred.	No.	0.0057:			
	Matches	395:	Conservative	0:	Mismatches	455:	Indels	12:	Gaps
									3:
QY	1559	ATATTATTGTTAGTTGGAAATTTAATAAGAGGAGCAACTTGAACATTACATAATTATATT	1618						
DB	324686	ATATTGTTAGTATCTTTTATTAATATACCAATATATTTTANCAATCAATTAATATTGGCAAC	324745						
QY	1619	AGATCTCTAGTATGTATATATCCCAATACATACCTTGGATGTTAAACCTTAATCTGTTT	1678						

Db 670 AAAAAATTTTATTAATNNAAAAAANANNTGNTAAATTTTAAATTTTTTTTTT 729

QY 2187 TTTTATGACAGCAAGAAATACAGATAGAAAATTTGGTGGAGATGAGATAG 2246

Db 730 TTTTAAAAAATTTAATTAATTTAATTTTAAAAAATTTTAAAAAATTTTGG 789

QY 2247 AATATACATTAAGCAAAATTTTAAAAATTAATATACCTATA 2288

Db 790 TTTAATTAATTAATTAATTAATTTTNNNTTTTAAAA 831

RESULT 6

US-10-239-676-18

Sequence 18, Application US/10239676

Publication No. US20030082609A1

GENERAL INFORMATION:

APPLICANT: OLEK, Alexander

APPLICANT: PIEPENBROCK, Christian

APPLICANT: BERLIN, Kurt

TITLE OF INVENTION: Diagnosis of Diseases Associated with Gene Regulation

FILE REFERENCE: 5013.1003

CURRENT APPLICATION NUMBER: US/10/239,676

CURRENT FILING DATE: 2002-09-24

PRIOR APPLICATION NUMBER: PCT/EP01/03968

DE 10019058.8

DE 10019173.8

DE 10032529.7

DE 10043826.1

PRIOR FILING DATE: 2001-04-06

2000-04-06

2000-04-07

2000-06-30

2000-09-01

NUMBER OF SEQ ID NOS: 228

SEQ ID NO 18

LENGTH: 5979

TYPE: DNA

ORGANISM: Artificial Sequence

FEATURE:

OTHER INFORMATION: chemically treated genomic DNA (Homo sapiens)

US-10-239-676-18

Query Match

Best Local Similarity 2.3%; Score 75.2; DB 9; Length 5979;

Matches 388; Conservative 0; Mismatches 468; Indels 12; Gaps 2;

QY 1554 TTACATATTTATTTAGTGTAGTATTAATTAAGCGCACTGTAACATTAATTT 1613

Db 1459 TAAAGTTAATGTATTAATTTTATTTAGAAATTTATTTTGTGTTTTTATTA 1518

QY 1614 AATATGATAGTATGATGTATTAATTAATTAATTAATTAATTAATTAATCT 1673

Db 1519 AATTTTAAATTTATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 1578

QY 1674 TGTTCCTGCTGCTATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 1733

Db 1579 TTATATTAATTAATTAATTAATTTTATTAATTAATTAATTAATTAATTAAT 1638

QY 1734 GCGATGATGAGATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 1793

Db 1639 TAAATGATGATTAATTTTAAATTAATTAATTAATTAATTAATTAATTAAT 1698

QY 1794 ATAGTTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 1853

Db 1699 AATATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 1758

QY 1854 AATTTTCTGCAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 1913

Db 1759 ATGTATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 1818

QY 1914 GAGTATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 1973

Db 1819 TTTTAAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 1878

QY 1974 TTATTA - AATATGATGCTATTAATTAATTAATTAATTAATTAATTAAT 2031

Db 1879 AATTTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 1938

QY 2032 AATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 2091

Db 1939 TTTTAAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 1988

QY 2092 TATTTTGAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 2151

Db 1989 AATTTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 2048

QY 2152 AAAAAATTTTGTAAACAGAACTCCAAATTTTAAAAATTAATTAATTAATTAATTAAT 2211

Db 2049 AATTTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 2108

QY 2212 GATGAAACATTAATTTGTTGAGATGAGATTAATTAATTAATTAATTAATTAATTAAT 2271

Db 2109 TATTTAATTAATTAATTTTGAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 2168

QY 2272 AATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 2331

Db 2169 AATGTTGCTGCTATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 2228

QY 2332 TGTGCGATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 2391

Db 2229 TAGTATGATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 2288

QY 2392 AATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 2419

Db 2289 AATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 2316

RESULT 7

US-10-239-676-90

Sequence 90, Application US/10239676

Publication No. US20030082609A1

GENERAL INFORMATION:

APPLICANT: OLEK, Alexander

APPLICANT: PIEPENBROCK, Christian

APPLICANT: BERLIN, Kurt

TITLE OF INVENTION: Diagnosis of Diseases Associated with Gene Regulation

FILE REFERENCE: 5013.1003

CURRENT APPLICATION NUMBER: US/10/239,676

CURRENT FILING DATE: 2002-09-24

PRIOR APPLICATION NUMBER: PCT/EP01/03968

DE 10019058.8

DE 10019173.8

DE 10032529.7

DE 10043826.1

PRIOR FILING DATE: 2001-04-06

2000-04-06

2000-04-07

2000-06-30

2000-09-01

NUMBER OF SEQ ID NOS: 228

SEQ ID NO 90

LENGTH: 5689

TYPE: DNA

ORGANISM: Artificial Sequence

FEATURE:

OTHER INFORMATION: chemically treated genomic DNA (Homo sapiens)

NAME/KEY: unsure

LOCATION: (1818, 1930, 1931, 1934, 1947, 1948, 1966)

US-10-239-676-90

Query Match

Best Local Similarity 2.3%; Score 73.4; DB 9; Length 5689;

Matches 304; Conservative 0; Mismatches 374; Indels 1; Gaps 1;

QY 1518 ATCTGACATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 1577


```

Db      5453 ATGGAAATTAATGATAGAAATTTAAATGAAATGAAATGAAAGAAAGAAAGAAATG 5511
QY      1591 GAACCTGTACATTAATCAATATTTATATAGTACTAGTATGATATTTCCAAATACATA 1650
Db      5512 TAAATATATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 5571
QY      1651 CTTTGAGTGTAACTTAATCTGTTCTCCACGCTAATTAATTAATTAATTAATTAATTAATTA 1709
Db      5572 GAATATTTGAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 5631
QY      1710 -----TAAAGAAAGTTTGTCTATTTCCGAGATGAAAGAAAGAAAGAAAGAAAGAAAG 1761
Db      5632 ATAAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 5691
QY      1762 TTTAAATTTTGAAGAAATGTAACCCCTTTACTAGTATGATTAATTAATTAATTAATTAATTA 1821
Db      5692 ATAGTATTTATGATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 5751
QY      1822 CCATTAATGACAGCCTTACAACTGTAGTACATTTTCTGCAAAATTAATTAATTAATTA 1881
Db      5752 AGTTTATGATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 5810
QY      1882 ATTCATATGCTACTATCAATGAAAGAAACAGCTGAGTATTAATTAATTAATTAATTAATTA 1941
Db      5811 TTTAAATTTTAAAGTTTAAATTTTAAATTTGAAAGGAGAAATTTTAAATTTATTTGT 5870
QY      1942 ATTTTGAAGAAATGTTATTTTCTAACATATTAATTAATTAATTAATTAATTAATTAATTA 2001
Db      5871 ATGATTTAGTATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 5930
QY      2002 TTCCTATGTTCTTAAATTAATTTTAAATTAATTAATTAATTAATTAATTAATTAATTA 2061
Db      5931 AAATTAATTTTAAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 5990
QY      2062 ATAGTGTGTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 2121
Db      5991 TTTATATAGTACGTTAAATTTATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 6050
QY      2122 TTATGCAATTAACATGCAATGAAAGTCCAAAGAAATTTGTTAACAGAACTTCCAA 2181
Db      6051 TAATGATGCTTTATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 6110
QY      2182 TTTTATTTTATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 6170
Db      6111 TTTTATTTTATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 6170
QY      2240 AATGATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 2281
Db      6171 ATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 6212

```

RESULT 10

```

US-10-239-676-35
; Sequence 35, Application US/10239676
; Publication No. US20030082609A1
; GENERAL INFORMATION:
; APPLICANT: OLEK, Alexander
; APPLICANT: PIEPERBROCK, Christian
; APPLICANT: BERLIN, Kurt
; TITLE OF INVENTION: Diagnosis of Diseases Associated with Gene Regulation
; FILE REFERENCE: 3013.1003
; CURRENT APPLICATION NUMBER: US/10/239, 676
; CURRENT FILING DATE: 2002-09-24
; PRIORITY APPLICATION NUMBER: PCT/EP01/03968
; DE 10019058.8
; DE 10019173.8
; DE 10032529.7
; DE 10043826.1
; PRIORITY FILING DATE: 2001-04-06
; 2000-04-06
; 2000-04-07
; 2000-06-30

```

```

; 2000-09-01
; NUMBER OF SEQ ID NOS: 228
; SEQ ID NO 35
; LENGTH: 12405
; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: chemically treated genomic DNA (Homo sapiens)
; NAME/KEY: unsure
; LOCATION: (7895)
; US-10-239-676-35

```

```

Query Match      2.2%; Score 71; DB 9; Length 12405;
Best Local Similarity 45.0%; Pred. No. 0.036;
Matches 347; Conservative 0; Mismatches 420; Indels 4; Gaps 2;

```

```

QY      1512 GTAACATCTGCACATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 1571
Db      190 GAGCCGTTTCCGTTTAAAGTAAGTAAATTAATTAATTAATTAATTAATTAATTAATTAATTA 249
QY      1572 GTTGCATTTAATTAAGCCAGCTGTAACATTAATTAATTAATTAATTAATTAATTAATTA 1631
Db      250 TTGCTAGTGTATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 309
QY      1632 TGATTTATCCAA--ATACATACCTTGGATGTTTAACTTAATTAATTAATTAATTAATTAATTA 1689
Db      310 TGTGTTGTTTATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 369
QY      1690 ATTAATTTATATCAATGAGTAAGTAAGTAAGTAAGTAAGTAAGTAAGTAAGTAAGTAAGTA 1749
Db      370 TTAGGTATTTATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 429
QY      1750 AAACCTATGACCTTAATTTTGAAGAAATGTAACCCCTTTACATGATTAATTAATTAATTA 1809
Db      430 TTTAGATATGATTAATTTTGTGTTATTTGTTATTTATTTATTAATTAATTAATTAATTAATTA 489
QY      1810 ATGTTTTTGTGCAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 1869
Db      490 AAGATTTTGTGTTTATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 549
QY      1870 ATTAATTTAGCAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 1927
Db      550 TTGTTATTTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 609
QY      1928 AATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 1987
Db      610 TAAATTTATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 669
QY      1988 TGCCATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 2047
Db      670 TGTATTTATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 729
QY      2048 ATTAATGACCAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 2107
Db      730 ATTTGTTTATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 789
QY      2108 CAATTTATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 2167
Db      790 GTTTTGTAAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 849
QY      2168 CAGAACTTCCAAATTTTATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 2227
Db      850 ATTTGATTTGAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 909
QY      2228 GTTGTGAATGAGATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 2278
Db      910 GTGTTTAAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 960

```

RESULT 11

```

US-09-754-853A-2/c
; Sequence 2, Application US/09754853A

```

```

; Publication No. US20030005491A1
; GENERAL INFORMATION:
; APPLICANT: Hauge, Brian M.
; APPLICANT: Parnell, Laurence D.
; APPLICANT: Parsons, Jeremy D.
; APPLICANT: Wang, Ming Li
; TITLE OF INVENTION: Nucleic Acid Molecules And Other Molecules Associated with
; FILE REFERENCE: 38-10(15810)B
; CURRENT APPLICATION NUMBER: US/09/754,853A
; PRIOR FILING DATE: 2001-01-05
; PRIOR APPLICATION NUMBER: US 60/174,880
; NUMBER OF SEQ ID NOS: 1119
; SEQ ID NO 2
; LENGTH: 335913
; TYPE: DNA
; ORGANISM: glycine max
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (45163)..(45314)..(45450)..(45509)..(46941)..(48763)..(48975)..(49573)
; OTHER INFORMATION: Clone ID: 240017_region_63
; US-09-754-853A-2

```

```

Query Match      2.1%; Score 68.8; DB 9; Length 335913;
Best Local Similarity 45.9%; Pred. No. 0.3;
Matches 390; Conservative 0; Mismatches 447; Indels 13; Gaps 4;

```

```

QY 1550 ATCTTAACATATTTATTTAGTTGGAATTTAAATAGAGCAACTGTGACATTACAT 1609
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 99512 ATTTGTATGTATTTTATATGATTTTAAATATGTAATATTTTATTTTAAAT 99453
QY 1610 ATTTATATAGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1669
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 99453 ATTTATATATTTTATATATGTTTATTTTAAATATTTGTAATTTTAAATTTG 99396
QY 1670 ATCTGTTCTCTCTACGCTATATATATATATATATATATATATATATATATAT 1729
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 99395 AGATATATTTTAT-----ATATTTTAAATTTTGTATTTTAAATATATATATAT 99340
QY 1730 TTTCGGATGATGAGAGATTAACCTATGACTTTATTTTGAAGATGTAACCTTTT 1789
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 99339 CGTACAAATTAATAGCACTAAGATATATTTTAAATTTTAAATTTTAAATTTTAA 99280
QY 1790 ACTCATAGATTAATACGTAATGTTTGTGCAATTAAGACGCTTACAACTGTAT 1849
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 99279 TTGTCAATTTTATATATATATATATATATATATATATATATATATATATAT 99220
QY 1850 AGTCATATTTTCTGCAATATATTAATTAAGATTCATGCTACTATGAGAGAAAC 1909
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 99219 AATTTCTCATACAAATTAATTAATTTTCAATTAATTAATTAATTAATTAATTA 99160
QY 1910 AACTGAGATTAATTTTAAATTAAGCAAAATTTTGAAGATGTAATTTGTAAC 1969
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 99159 A---AAATATTTTAACTTTTAAATTTTAAATTTTAAATTTTAAATTTTAAAG 99103
QY 1970 AATATATTAATTAATNGATGCCATATATATTTTCTATGCTTAAATATTTT 2029
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 99102 AAACAAATTAATTAATCATATATATTTTATTTTATTAACAAACAAATCACT 99043
QY 2030 AATATATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 2089
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 99042 CATATACATTAATTAAGAAATTAATTAATTTTAAATTAACAAATTAATTAATTA 98983
QY 2090 AATATTTTGAATCTACAAAT---TATTAATTTTACATTAACATGCAATGAGAAG 2146
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 98982 AATTAATCTAAATTAATTAATTTTAAATTAATTAATTAATTAATTAATTAATTA 98923
QY 2147 TTCCAAAAAATTTTGTGAAGAGAACTCCAAATTTTATTTTATGGAACAGAA 2206
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 98922 TTTTAAAGAAATTAAGAAATTAATTAATTAATTAATTAATTAATTAATTAATTA 98863
QY 2207 TAACAGATAGAAAACTATTTTGTGTGAATGAGAGTATATATACATTAACAAATTT 2266

```

```

DB 98862 AAAATTTTAAACATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 98803
QY 2267 TAAAAATTAATTAAGCCATATCGGCTCAAGTATGTAATGATGATTAATTAATTA 2326
DB 98802 ACAGCTTAAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 98743
QY 2327 ATGATGTCGATTCAGATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 2386
DB 98742 ATGAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 98683
QY 2387 AATTAATTAATTA 2396
DB 98682 AAAATTAATTA 98673

```

RESULT 12

```

; US-09-754-853A-3/c
; Sequence 3, Application US/09754853A
; Publication No. US20030005491A1
; GENERAL INFORMATION:
; APPLICANT: Hauge, Brian M.
; APPLICANT: Parnell, Laurence D.
; APPLICANT: Parsons, Jeremy D.
; APPLICANT: Wang, Ming Li
; TITLE OF INVENTION: Nucleic Acid Molecules And Other Molecules Associated with
; FILE REFERENCE: 38-10(15810)B
; CURRENT APPLICATION NUMBER: US/09/754,853A
; PRIOR FILING DATE: 2001-01-05
; PRIOR APPLICATION NUMBER: US 60/174,880
; NUMBER OF SEQ ID NOS: 1119
; SEQ ID NO 3
; LENGTH: 335913
; TYPE: DNA
; ORGANISM: glycine max
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (46798)..(48763)..(48975)..(49573)
; OTHER INFORMATION: Clone ID: 240017_region_63
; US-09-754-853A-3

```

```

Query Match      2.1%; Score 68.8; DB 9; Length 335913;
Best Local Similarity 45.9%; Pred. No. 0.3;
Matches 390; Conservative 0; Mismatches 447; Indels 13; Gaps 4;

```

```

QY 1550 ATCTTAACATATTTATTTAGTTGGAATTTAAATAGAGCAACTGTGACATTACAT 1609
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 99512 ATTTGTATGTATTTTATATGATTTTAAATATGTAATATTTTATTTTAAAT 99453
QY 1610 ATTTATATAGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1669
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 99453 ATTTATATATTTTATATATGTTTATTTTAAATATTTGTAATTTTAAATTTG 99396
QY 1670 ATCTGTTCTCTCTACGCTATATATATATATATATATATATATATATATATAT 1729
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 99395 AGATATATTTTAT-----ATATTTTAAATTTTGTATTTTAAATTAATTA 99340
QY 1730 TTTCGGATGATGAGAGATTAACCTATGACTTTATTTTGAAGATGTAATTTGTAAC 1789
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 99339 CGTACAAATTAATAGCACTAAGATATATTTTAAATTTTAAATTTTAAATTTTAA 99280
QY 1790 ACTCATAGATTAATACGTAATGTTTGTGCAATTAAGACGCTTACAACTGTAT 1849
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 99279 TTGTCAATTTTATATATATATATATATATATATATATATATATATATATAT 99220
QY 1850 AGTCATATTTTCTGCAATATATTAATTAAGATTCATGCTACTATGAGAGAAAC 1909
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 99219 AATTTCTCATACAAATTAATTAATTTTCAATTAATTAATTAATTAATTAATTA 99160
QY 1910 AACTGAGATTAATTTTAAATTAAGCAAAATTTTGAAGATGTAATTTGTAAC 1969
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

```

Db 99159 A---AAATATTTTAACTTTTAAAAATTTAACTTTTAAAAATTTAAGATTAAAAAGA 99103
 Oy 1970 AATATTATTAAAAATGATGCCCTAATAAGTATTCCATGTTCTTAAAAATATTTTCTT 2029
 Db 99102 AAAACAAATATTAAATATCATTTATGTAAATTTTATTATTATTTACAAAAACAATCAACT 99043
 Oy 2030 AATATTAGTATATAAATTCATTATGACCAATTAATAGTGGTGAATTCAAATATCTCCATT 2089
 Db 99042 CAATACATTTATTGAAAAATTTATATTAATTTTAAAAATACAAAATATATTAAATGAACT 98983
 Oy 2090 AATATTTTGGAAATCTACAAAT--TATTAAATTTTAGTCATTAACAATGATGAGAAAG 2146
 Db 98982 AATATAGTAAATAATTTAAATATTATTATTTAAATTAATAAATAAATAATATTATTAAAT 98923
 Oy 2147 TTCCAAAAAAAATTTTGTACGAAACCTCCAAATTTTTTTATTGACAAAGAAA 2206
 Db 98922 TTTTAAAGAAATGAAAAATTTAAAGTATTAATAAAATTAACAATAAATAAATCATAAACA 98863
 Oy 2207 TAACAGATTAACAAAATCTTTTGTGTGGAATGGAAGTAGAATATATCATTAAGCAAAATT 2266
 Db 98862 AAAAATTTTAAACATATAAAATTAATATGATTTTAAAGTCATATAAAACAATAAAAATATT 98803
 Oy 2267 TAAAAAATTTATATAGCTTATACGCGCTCAAGATGATGTATCTAGTAGGTGTAATTATA 2326
 Db 98802 ACAGCTTTAAATATATCATAGTTAAATAAAAAAATAAACAATATATATTTTAAAGTCATA 98743
 Oy 2327 ATGATGTCGTCGATTCGATTTGGACACAAATGAAAACGGAATTAATTAATTACTTTA 2386
 Db 98742 ATGAAATTAATAAATTTAAACTGAATTCCTTAAAAAATAAATAGATTTCAGTTAAAAAATPA 98683
 Oy 2387 AAATTAATAA 2396
 Db 98682 AAAATATAA 98673

 RESULT 13
 US-10-239-676-64/C
 : Sequence 64, Application US/10239676
 : Publication No. US2003082609A1
 : GENERAL INFORMATION:
 : APPLICANT: OLEK, Alexander
 : APPLICANT: PIEPENBROCK, Christian
 : APPLICANT: BERLIN, Kurt
 : TITLE OF INVENTION: Diagnosis of Diseases Associated with Gene Regulation
 : FILE REFERENCE: 5013.1003
 : CURRENT APPLICATION NUMBER: US/10/239, 676
 : CURRENT FILING DATE: 2002-09-24
 : PRIOR APPLICATION NUMBER: PCT/EP01/03968
 : DE 10019058.8
 : DE 10019173.8
 : DE 10032529.7
 : DE 10043826.1
 : PRIOR FILING DATE: 2001-04-06
 : 2000-04-06
 : 2000-04-07
 : 2000-06-30
 : 2000-09-01
 : NUMBER OF SEQ ID NOS: 228
 : SEQ ID NO 64
 : LENGTH: 6298
 : TYPE: DNA
 : ORGANISM: Artificial Sequence
 : FEATURE:
 : OTHER INFORMATION: chemically treated genomic DNA (Homo sapiens)
 : US-10-239-676-64

Query Match	2.1%	Score 68;	DB 9;	Length 6258;
Best Local Similarity	48.8%;	Pred. No. 0.087;		
Matches 273;	Conservative 0;	Mismatches 280;	Indels 7;	Gaps 3

Oy
Db

1850 AGCAATTTTTCGCAAATATTAAATAGCAATCGAATCTACATAACAATAGAGAACC 1907
 | || | | | | | | |
4117 AATCACATTATTAATAATATATATAAAAAAAAAATTTAAAATTTTATCCGTATATAAAAATTT 4056

OY	1910	AGCGAGATTAACATTAAATTTTAAAGCAAAAATTTTTGAAAAAATGTATATACTTCAC	1969
Dd	4057	TCCITTTTAATATCATATTATTAA--AAATATTAAATTA AAAAATCATTATAT	4001
OY	1970	AATATTAATAAATATGATGCCCTAATATGATATTCCTATGCTCTTAAATATATTTTTTT	2029
Dd	4000	ACCTTTAATACATATAAATAATTAATTAATTAATTAATTAATTAATTAATTAATTT	3941
OY	2030	ATATTTAGTTTAATATACATATATACCAATAAATAGTTGGCAATTCAAAATATCCATT	2089
Dd	3940	ATATTTATATAACAATATATATATATATACATACACTTACTTCTTCATTTTCARAT	3881
OY	2090	AATATTTTTTGAAATCTACA-AAATATTATATTTTAGTCAATAACAATGCATAGAAGTT	2148
Dd	3880	TATATTATATATATACATATACATTAACAATATTTTATACATTTTAAAAAANAAT	3821
OY	2149	CCAAAAAATTTTTGTTAACAGAACTCCAAAATTTTTTTTTTTTAAAGAACGAATA	2208
Dd	3820	TAACTTAAAAACATATTTAATAAATATTCAAAAATTTTTTACTTAATCAAATTTTCATT-	3762
OY	2209	ACAGATGAAAACTATTTTGTGTGGATGAGAGACTAATATATACATTAAAGCAATTTTA	2268
Dd	3761	--CTACTTTAACTTTTAAAAATATCCA AAAAAAAAAATTTTAAATACTTAATTTCTATTA	3704
OY	2269	AAAAATATATAACCCATACAGCGCCCAAGTATAGTATCTAGTAGAGGTATATTAATAT	2328
Dd	3703	CAAAATACTTATCTTTTAAATATCACAAAACATTTAATATCTATATCACTAACCAAAAATBA	3644
OY	2329	GCAATGGTCGATTCAGATTTGGGACACAAATGAANAAGGAATTTAAATATTMACTTAAA	2388
Dd	3643	TTACTATTAATATATTA AAAAAAAAAAATAAATAAATAAATATTAATAAATCTCTACAACTATCTCATTAAT	3584
OY	2389	ATAAATAAATAATTGAGTAA 2408	
Dd	3583	AAAAATATAAACACGATBAA 3564	

```

RESULT 14
US-10-239-676-20/c
; Sequence 20, Application US/10239676
; Publication No. US20030082609A1
; GENERAL INFORMATION:
; APPLICANT: OLEK, Alexander
; APPLICANT: PIEPENBROCK, Christian
; APPLICANT: BERLIN, Kurt
; TITLE OF INVENTION: Diagnosis of Diseases Associated with Gene Regulation
; FILE REFERENCE: 5013.1003
; CURRENT APPLICATION NUMBER: US/10/239.676
; CURRENT FILING DATE: 2002-09-24
; PRIOR APPLICATION NUMBER: PCT/EP01/03968
; DE 10019058.8
; DE 10019173.8
; DE 10032529.7
; DE 10043826.1
; PRIOR FILING DATE: 2001-04-06
; 2000-04-06
; 2000-04-07
; 2000-06-30
; 2000-09-01
; NUMBER OF SEQ ID NOS: 228
; SEQ ID NO 20
; LENGTH: 11260
; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: chemically treated genomic DNA (Homo sapiens)
US-10-239-676-20

```

Query Match	2.1%	Score 68;	DB 9;	Length 11260;
Best Local Similarity	46.4%	Pred. No. 0.11;		
Matches 256;	Conservative 0;	Mismatches 295;	Indels 1;	Gaps 1

QY 1839 ACAACGCTGGTATTCATTTTCTTTCGGAATATTTAAATTTGAATTCGAATGCTATATCA 1898
 Db 6329 AAATATTTCAAAACGATTTTCTTAAACATATATAAACCACTTTATTTTTCATTAATTCGTACCA 6270
 QY 1899 ATAGAAGAAACAGCTGAGTATTCATTTTATTTAAAGACAAATTTTGAAAATGCTTA 1958
 Db 6269 TAAAAAATAAATATATATTAATATATTCACCTTATTTAAATTCGAATATATTAATTCGA 6210
 QY 1959 TAATTTCTACCAATATATTTAAATATATGATGCTATATATTTTCCATGTTCTTAAAA 2018
 Db 6209 CCAAAATATATACCAATATTTTATATTTTAAACCACTTTTTCATCTTTTAATATATAAAA 6150
 QY 2019 TATTTTCTTTTATTTAGTATTTAATTAATACATTAATGACCAATATAGTTGGTGAATTCAA 2078
 Db 6149 TCTTTTATACCTATTTTCTATTTAAATATATTTTATAT -ATTTTCTTCATACAAAACAA 6091
 QY 2079 ATATCTCATTAATATTTTGTGAAATCTCAAAATTTTATATTTTATTTAGTCATATACAAATGC 2138
 Db 6090 ATCATCTACTTATATATATATATTTAACTAATATACCCATTTTATTTTAAATATCAATTT 6031
 QY 2139 ATAGAAGAGTCCAAAAAAATTTTGTTAACAGAACTCCAAATTTTCTTTTATTTGA 2198
 Db 6030 AAAAAACATTAATACACTTAACCTATTAACAAACATTAACATATTTTAAATTTTAAATTC 5971
 QY 2199 ACAGAAGATACAGATAGAAAACCTATTTTGTGTGAAATGGAAGTATATACATTTA 2258
 Db 5970 CTCTAAATATATTTCTATCAATCCCATTAATATTTCTATATAAAAATATTCGAATATACA 5911
 QY 2259 GCAAAATTTTAAAAATATATATATAGCCTATATCGCGCTCAAGATATGTTATCTAGTAGGTGT 2318
 Db 5910 ATAACTTTAAACAAAATATACCATATCTTAATATCTCCCTAATCATTTTACATTAACATAT -5851
 QY 2319 AATTATATATGATGATGGATTCAGAAATTTGGACACATGAAAACGAATTTAAATAT 2378
 Db 5850 TAAAAACATATAAAAATCTACCAAAAAAATTTTAAAAAATTTTAAAAAATTTTAAAAAT 5791
 QY 2379 TAACTTTTAAAT 2390
 Db 5790 ATATTTTAAACAT 5779

RESULT 15
 US-10-087-464-9/c
 : Sequence 9, Application US/10087464
 : Publication No. US20030059436a1
 : GENERAL INFORMATION:
 : APPLICANT: Chisholm, Althar
 : APPLICANT: Oh, Steven
 : APPLICANT: Liu, David
 : APPLICANT: Goel, Vikas
 : APPLICANT: Li, Xuerong
 : TITLE OF INVENTION: Band 3 Antigenic Peptides, Malaria Polypeptides and Uses Thereof
 : CURRENT APPLICATION NUMBER: US/10/087,464
 : PRIOR FILING DATE: 2002-03-01
 : NUMBER OF SEQ ID NOS: 59
 : SOFTWARE: PatentIn version 3.0
 : SEQ ID NO 9
 : LENGTH: 5917
 : TYPE: DNA
 : ORGANISM: Plasmodium falciparum
 : US-10-087-464-9

Query Match	2.1%;	Score 67.4;	DB 9;	Length 5917;
Best Local Similarity	48.7%;	Pred. No. 0.11;		
Matches 211;	Conservative 0;	Mismatches 221;	Indels 1;	Gaps 1

D

QY 1848 5867	<div>ATAGCAATTCTTCGCAAAATAAATAGCATGTACTATACATAGACA</div> <div> - </div> <div>AAAAAAAATTATGTTATTAAAGTATTTAATTTGCAGTGTTTAAAA</div>	1907 5808
------------------------	--	--------------

QY	1908	ACAGCTGGATATTCACATTTTAAATTTAAAGACAAAATTTTGAAAACTTATATTTCTA	1967
Dd	5807	AAAAAAAAATGCTTTTATGTATGCGTAAAAAAAATATATATATATATATATATATGTA	5748
QY	1968	ACAAATATTTTAAATATGATGCGCTATATATGATATTTCCATGTTCTTAAATATATTTT	2027
Dd	5747	AAAT	5688
QY	2028	TTATATTTAGTTATTAATACATTATGAAACCAATATAGTTGGTGAATTCAAATATCTCA	2087
Dd	5687	ATATATTTAAATTTACGACTTTTATTTTATTTTAAAAAGTTTATTTT - AAATCTATATA	5629
QY	2088	TTATATTTTGGAAATCTCAAAATTTTATATATTTAGTCATATACATATGATGAAGT	2147
Dd	5628	TAAATATTTAAATATGAACCTGAAAAAGAAAGAAAAAAAATATATATATATATATATGTA	5569
QY	2148	TCCAAAAAAATTTTGTGAACAGAACTTCCAAATTTTATTTTATTTTATGGAACAGAAAT	2207
Dd	5568	TAAATGTATTAATTTTATAGTATATATATTTTATTTTATTTTATTTTATATATATATTTT	5509
QY	2208	AACAGATAGAAAATCTATTTTGTGTGGAATGGAAGTATATATATATATATAGCAAAATTTT	2267
Dd	5508	ATATATTTTAAATATATTTTGAACATAGCTAAATATATCAGTATATCTGTATATATCTAAAT	5449
QY	2268	AAAAAATTAATA 2280	
Dd	5448	ATATATATATATA 5436	

Search completed: June 11, 2003, 08:07:41
Job time : 463.588 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: June 10, 2003, 22:03:39 ; Search time 7692.77 Seconds

(without alignments)
10465.184 Million cell updates/sec

Title: US-09-502-426b-1_COPY_1_3202

Perfect score: 3202
Sequence: 1 atgtgtgtatattatgtt.....gagagagagaactagctc 3202

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 24791104 seqs, 12571243825 residues

Total number of hits satisfying chosen parameters: 49582208

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%

Maximum Match 100%
Listing first 45 summaries

Database :

```

Pending_Patents_NA_Main:*
1: /cgn2_6/ptodata/1/pna/US0600.COMB.seq:*
2: /cgn2_6/ptodata/1/pna/US06.COMB.seq:*
3: /cgn2_6/ptodata/1/pna/US07.COMB.seq:*
4: /cgn2_6/ptodata/1/pna/US08.COMB.seq:*
5: /cgn2_6/ptodata/1/pna/US081.COMB.seq:*
6: /cgn2_6/ptodata/1/pna/US082.COMB.seq:*
7: /cgn2_6/ptodata/1/pna/US083.COMB.seq:*
8: /cgn2_6/ptodata/1/pna/US084.COMB.seq:*
9: /cgn2_6/ptodata/1/pna/US085.COMB.seq:*
10: /cgn2_6/ptodata/1/pna/US086.COMB.seq:*
11: /cgn2_6/ptodata/1/pna/US087.COMB.seq:*
12: /cgn2_6/ptodata/1/pna/US088.COMB.seq:*
13: /cgn2_6/ptodata/1/pna/US089.COMB.seq:*
14: /cgn2_6/ptodata/1/pna/US090.COMB.seq:*
15: /cgn2_6/ptodata/1/pna/US091.COMB.seq:*
16: /cgn2_6/ptodata/1/pna/US092.COMB.seq:*
17: /cgn2_6/ptodata/1/pna/US093.COMB.seq:*
18: /cgn2_6/ptodata/1/pna/US094.COMB.seq:*
19: /cgn2_6/ptodata/1/pna/US095A.COMB.seq:*
20: /cgn2_6/ptodata/1/pna/US095B.COMB.seq:*
21: /cgn2_6/ptodata/1/pna/US095C.COMB.seq:*
22: /cgn2_6/ptodata/1/pna/US095D.COMB.seq:*
23: /cgn2_6/ptodata/1/pna/US096A.COMB.seq:*
24: /cgn2_6/ptodata/1/pna/US096B.COMB.seq:*
25: /cgn2_6/ptodata/1/pna/US096C.COMB.seq:*
26: /cgn2_6/ptodata/1/pna/US096D.COMB.seq:*
27: /cgn2_6/ptodata/1/pna/US096E.COMB.seq:*
28: /cgn2_6/ptodata/1/pna/US097A.COMB.seq:*
29: /cgn2_6/ptodata/1/pna/US097C.COMB.seq:*
30: /cgn2_6/ptodata/1/pna/US097C.COMB.seq:*
31: /cgn2_6/ptodata/1/pna/US098A.COMB.seq:*
32: /cgn2_6/ptodata/1/pna/US098B.COMB.seq:*
33: /cgn2_6/ptodata/1/pna/US098C.COMB.seq:*
34: /cgn2_6/ptodata/1/pna/US099A.COMB.seq:*
35: /cgn2_6/ptodata/1/pna/US099C.COMB.seq:*
36: /cgn2_6/ptodata/1/pna/US099C.COMB.seq:*
37: /cgn2_6/ptodata/1/pna/US099D.COMB.seq:*
38: /cgn2_6/ptodata/1/pna/US100A.COMB.seq:*
39: /cgn2_6/ptodata/1/pna/US100B.COMB.seq:*
40: /cgn2_6/ptodata/1/pna/US101A.COMB.seq:*
41: /cgn2_6/ptodata/1/pna/US101B.COMB.seq:*
42: /cgn2_6/ptodata/1/pna/US102A.COMB.seq:*
43: /cgn2_6/ptodata/1/pna/US102B.COMB.seq:*

```

Result No.	Score	Query Match	Length	DB ID	Description
1	3202	100.0	6888	1	PCT-US500-03820-1
2	3202	100.0	6888	19	US-09-502-426-1
3	3134.2	97.9	84196	27	US-09-692-412-94
4	3134.2	97.9	84196	31	US-09-803-736-1074
5	960.4	30.0	1699	18	US-09-451-320-2050
6	125	3.9	826	1	PCT-US99-228538-3486
7	118	3.7	84196	27	US-09-692-412-94
8	118	3.7	84196	31	US-09-803-736-1074
9	105.6	3.3	1016	64	US-60-207-458-33653
10	102.4	3.2	1326	26	US-09-663-779-8205
11	98.2	3.1	1326	26	US-09-663-779-8205
12	97.6	3.0	810	26	US-09-666-355A-6400
13	97.6	3.0	22350	68	US-60-245-228-17
14	94.6	3.0	742	26	US-09-663-779-3024
15	91.6	2.9	960	41	US-10-198-846-6381
16	91.6	2.9	464387	35	US-09-948-941-643
17	89.2	2.8	98573	68	US-60-248-505-4421
18	88.4	2.8	1368	38	US-10-015-127-2421
19	87.6	2.7	908	56	US-60-126-265-2038
20	87.6	2.7	27978	64	US-60-207-423-22
21	87.6	2.7	32768	64	US-60-207-423-23

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

22 87.6 2.7 32768 64 US-60-207-905-22 Sequence 22, Appl
23 87.6 2.7 32768 64 US-60-207-905-23 Sequence 23, Appl
24 87.6 2.7 32768 64 US-60-207-905-24 Sequence 24, Appl
25 87.6 2.7 32768 64 US-60-208-837-26 Sequence 26, Appl
26 87.6 2.7 32768 64 US-60-208-837-27 Sequence 27, Appl
27 87.6 2.7 32768 64 US-60-208-837-28 Sequence 28, Appl
28 87.4 2.7 923 64 US-60-207-458-34124 Sequence 34124, A
29 87.4 2.7 28034 65 US-60-212-657-88 Sequence 88, Appl
30 87.4 2.7 32768 65 US-60-213-181-33 Sequence 33, Appl
31 87.4 2.7 32768 65 US-60-213-181-34 Sequence 34, Appl
32 87.4 2.7 32768 65 US-60-213-181-35 Sequence 35, Appl
33 87.4 2.7 158001 42 US-10-211-179-11 GENERAL INFO:MART
34 87 2.7 868 25 US-09-654-617-367906 Sequence 367906,
35 87 2.7 868 27 US-09-684-016-367906 Sequence 367906,
36 87 2.7 1368 38 US-10-015-127-2421 Sequence 2421, App
37 87 2.7 192014 68 US-60-248-823-33 Sequence 33, Appl
38 87 2.7 755 24 US-09-634-3068-128109 Sequence 128109,
39 86.8 2.7 755 38 US-10-027-632-128109 Sequence 128109,
40 86.6 2.7 720 18 US-09-404-520-12853 Sequence 12853, A
41 86.6 2.7 961 56 US-60-126-265-3039 Sequence 3039, A
42 86.6 2.7 32768 65 US-60-213-177-227 Sequence 227, App
43 86.4 2.7 38678 35 US-09-948-947-136 Sequence 136, App
44 86.4 2.7 38684 35 US-09-948-947-154 Sequence 154, App
45 86.4 2.7 38684 35 US-09-948-947-154 Sequence 154, App

ALIGNMENTS

RESULT 1
PCT-US00-03820-1
Sequence 1, Application PC/TUS0003820
GENERAL INFORMATION:
APPLICANT: THE ARIZONA BOARD OF REGENTS ON BEHALF OF THE UNIVERSITY OF ARIZONA.
TITLE OF INVENTION: DMF4 POLYNUCLEOTIDES, POLYPEPTIDES AND USES THEREOF
FILE REFERENCE: 2225-0001 40
CURRENT APPLICATION NUMBER: PCT/US00/03820
CURRENT FILING DATE: 2000-02-11
EARLIER APPLICATION NUMBER: 60/119,657
EARLIER FILING DATE: 1999-02-11
EARLIER APPLICATION NUMBER: 60/119,658
EARLIER FILING DATE: 1999-02-11
NUMBER OF SEQ ID NOS: 18
SOFTWARE: PatentIn Ver. 2.0
SEQ ID NO 1
LENGTH: 6888
TYPE: DNA
ORGANISM: Arabidopsis sp.
PCT-US00-03820-1

Query Match 100.0%; Score 3202; DB 1; Length 6888;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 3202; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 ATGTGGTATATATATGTTGGTTCGTTGACGTACATATATAATTTCTGTTCTGCT 60
DB 1 ATGTGGTATATATATGTTGGTTCGTTGACGTACATATATAATTTCTGTTCTGCT 60
QY 61 TATTCGTTACAGATTTGAGTTGGTCTCAATTTGATTCACAGATATTAATAT 120
DB 61 TATTCGTTACAGATTTGAGTTGGTCTCAATTTGATTCACAGATATTAATAT 120
QY 121 AAAATTCATTTAAATATATTAACAAGTAATTAATTAATTAATTAATTAATTA 180
DB 121 AAAATTCATTTAAATATATTAACAAGTAATTAATTAATTAATTAATTAATTA 180
QY 181 AATATCATCTTGGTATATGAGAAATATGAGTTTGAATTTATATATTAAGGAAA 240
DB 181 AATATCATCTTGGTATATGAGAAATATGAGTTTGAATTTATATATTAAGGAAA 240
QY 241 TAATGATTCATTTGTTGGATTTACACAGTTAAGTTTGTGTTCTTTGTTATATGT 300
DB 241 TAATGATTCATTTGTTGGATTTACACAGTTAAGTTTGTGTTCTTTGTTATATGT 300

QY 301 ATATGACTAAATCAAAAAAGATATGATTAAGATTAACATATTTCTGTTATGACCCCA 360
DB 301 ATATGACTAAATCAAAAAAGATATGATTAAGATTAACATATTTCTGTTATGACCCCA 360
QY 361 AAAAAAAGAAAAAAGAAAAAAGAAAAAAGAAAAAAGAAAAAAGAAAAAAGAAAAA 420
DB 361 AAAAAAAGAAAAAAGAAAAAAGAAAAAAGAAAAAAGAAAAAAGAAAAAAGAAAAA 420
QY 421 AGGTTATTTGATCATATTAATGATGATGATTTCTGTTGATTAATTAATTAATTA 480
DB 421 AGGTTATTTGATCATATTAATGATGATGATTTCTGTTGATTAATTAATTAATTA 480
QY 481 CAATTAATTTGCAATTCATATCTGTTGATTAATTAATTAATTAATTAATTAATTA 540
DB 481 CAATTAATTTGCAATTCATATCTGTTGATTAATTAATTAATTAATTAATTAATTA 540
QY 541 GTTTATGATCAGTCCCAATGATGATGATTTCTGTTGATTAATTAATTAATTAATTA 600
DB 541 GTTTATGATCAGTCCCAATGATGATGATTTCTGTTGATTAATTAATTAATTAATTA 600
QY 601 AGAGTATTTGAGAAAAAAGAAAAAAGAAAAAAGAAAAAAGAAAAAAGAAAAAAG 660
DB 601 AGAGTATTTGAGAAAAAAGAAAAAAGAAAAAAGAAAAAAGAAAAAAGAAAAAAG 660
QY 661 TGAATTAATTTCAAGAGAAATTAAGAGATGACAAACAAAGGTTGGAATTAATGCTCT 720
DB 661 TGAATTAATTTCAAGAGAAATTAAGAGATGACAAACAAAGGTTGGAATTAATGCTCT 720
QY 721 GCCAGCTTCTCACAATCAATATGACACCTATTGATTTCTGATATTTGTTAAA 780
DB 721 GCCAGCTTCTCACAATCAATATGACACCTATTGATTTCTGATATTTGTTAAA 780
QY 781 TTGCGATTAAGATGAGAAATTAATTTATTTGATGATGATGATGATGATGATGAT 840
DB 781 TTGCGATTAAGATGAGAAATTAATTTATTTGATGATGATGATGATGATGATGAT 840
QY 841 GGTATTTGCAATATCTGTTTAAACATATTTGCTGTTGTTGTTGTTGTTGTTGTT 900
DB 841 GGTATTTGCAATATCTGTTTAAACATATTTGCTGTTGTTGTTGTTGTTGTTGTT 900
QY 901 TAATCATATTTATGCGGATATATGATTAACATATGATATGATATGATATGATATG 960
DB 901 TAATCATATTTATGCGGATATATGATTAACATATGATATGATATGATATGATATG 960
QY 961 ACCATTTGATTAATCTTTTCAACATTTAGGAGCTGAGCTGAGCTGAGCTGAGCT 1020
DB 961 ACCATTTGATTAATCTTTTCAACATTTAGGAGCTGAGCTGAGCTGAGCTGAGCT 1020
QY 1021 CGATTTTACAGCTCACTAGTTAGATTAACATTAAGCATTAAGGACCCGTTCAAG 1080
DB 1021 CGATTTTACAGCTCACTAGTTAGATTAACATTAAGCATTAAGGACCCGTTCAAG 1080
QY 1081 TATTTATACAAAGTATACAACTGATATGATTAATTTCTGTTGATTAATTAATTA 1140
DB 1081 TATTTATACAAAGTATACAACTGATATGATTAATTTCTGTTGATTAATTAATTA 1140
QY 1141 CCGGTTGATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 1200
DB 1141 CCGGTTGATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 1200
QY 1201 ACATATTAACATTAATTTCTTACAGAAAAAACAACCTTAAGAGAAATTAACATTAAT 1260
DB 1201 ACATATTAACATTAATTTCTTACAGAAAAAACAACCTTAAGAGAAATTAACATTAAT 1260
QY 1261 GGGTATGCTATACCTTTACAGATATGATATGATTAATTAATTAATTAATTAATTA 1320
DB 1261 GGGTATGCTATACCTTTACAGATATGATATGATTAATTAATTAATTAATTAATTA 1320
QY 1321 GATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 1380
DB 1321 GATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 1380

1381 ATATCATCTCTGGTGGTCAACATTAACAAACAAACAAACAAACAAACAAACATTTT 1440
1381 ATATCATCTCTGGTGGTGGTCAACATTAACAAACAAACAAACAAACAAACAAACATTTT 1440
1441 CTGGATTCATTCATTAATGATCTAAATAGCATATCTTTGGGTTACAGTTTCAGATCC 1500
1441 CTGGATTCATTCATTAATGATCTAAATAGCATATCTTTGGGTTACAGTTTCAGATCC 1500
1501 TCATCAAGCGGTGTAACCATCTGCAACTATTAATGCTTTCTTTAATGATCTTTAAT 1560
1501 TCATCAAGCGGTGTAACCATCTGCAACTATTAATGCTTTCTTTAATGATCTTTAAT 1560
1561 ATTTATGTTAGTTGGAATTTAATAGAGCAACTGTATACATTAATTAATTAATAG 1620
1561 ATTTATGTTAGTTGGAATTTAATAGAGCAACTGTATACATTAATTAATTAATAG 1620
1621 ATATCATCTCTGGTGGTGGTCAACATTAATGATCTTTGGGTTACAGTTTCAGATCC 1680
1621 ATATCATCTCTGGTGGTGGTGGTCAACATTAATGATCTTTGGGTTACAGTTTCAGATCC 1680
1681 TCCTACGGTATTAATTAATTAATCAATGAGTAAACAAAGTTTGTCTTAATTTGGCATGC 1740
1681 TCCTACGGTATTAATTAATTAATCAATGAGTAAACAAAGTTTGTCTTAATTTGGCATGC 1740
1741 ATGAGGATTAACCTTAATGATCTTTAATTTTGAATAGTAAACCTTTTACTCATAGAT 1800
1741 ATGAGGATTAACCTTAATGATCTTTAATTTTGAATAGTAAACCTTTTACTCATAGAT 1800
1801 AATATCCGATATGTTTGTGGTGGTCAACATTAATGATCTTTGGGTTACAGTTTCAGATCC 1860
1801 AATATCCGATATGTTTGTGGTGGTGGTCAACATTAATGATCTTTGGGTTACAGTTTCAGATCC 1860
1861 TCATCAAGCGGTGTAACCATCTGCAACTATTAATGCTTTCTTTAATGATCTTTAAT 1920
1861 TCATCAAGCGGTGTAACCATCTGCAACTATTAATGCTTTCTTTAATGATCTTTAAT 1920
1921 ACATTTAATTTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 1980
1921 ACATTTAATTTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 1980
1981 AATATGATCCGATATGTTTGTGGTGGTCAACATTAATGATCTTTGGGTTACAGTTTCAGATCC 2040
1981 AATATGATCCGATATGTTTGTGGTGGTGGTCAACATTAATGATCTTTGGGTTACAGTTTCAGATCC 2040
2041 TAAATCATTAATGAGCAATTAATGATCTTTAATGATCTTTAATGATCTTTAATGATCTTT 2100
2041 TAAATCATTAATGAGCAATTAATGATCTTTAATGATCTTTAATGATCTTTAATGATCTTT 2100
2101 AATATGATCCGATATGTTTGTGGTGGTCAACATTAATGATCTTTGGGTTACAGTTTCAGATCC 2160
2101 AATATGATCCGATATGTTTGTGGTGGTGGTCAACATTAATGATCTTTGGGTTACAGTTTCAGATCC 2160
2161 TGTGTAACGAAACCTCCAAATTTTATGAGCAACAAAGATTAACAGATTAAGAAA 2220
2161 TGTGTAACGAAACCTCCAAATTTTATGAGCAACAAAGATTAACAGATTAAGAAA 2220
2221 CATATTTGTTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGG 2280
2221 CATATTTGTTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGG 2280
2281 AGCCTATACGCGCTCAAGATGATATCTAGTATGATTAATTAATTAATTAATTAATTAAT 2340
2281 AGCCTATACGCGCTCAAGATGATATCTAGTATGATTAATTAATTAATTAATTAATTAAT 2340
2341 TCAGATTAATGAGCAATTAATGATCTTTAATGATCTTTAATGATCTTTAATGATCTTT 2400
2341 TCAGATTAATGAGCAATTAATGATCTTTAATGATCTTTAATGATCTTTAATGATCTTT 2400
2401 TTGAGTAAATGTTTGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGG 2460
2401 TTGAGTAAATGTTTGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGG 2460
2461 GGGTTTGCATCTCCAGTTGGTAAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 2520

2461 GGGTTTGCATCTCCAGTTGGTAAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 2520
2521 TAGGGTCCCTTGCATATTTTCACTGCTTACCCCTACGCGAGCCGCTTTTCCCA 2580
2521 TAGGGTCCCTTGCATATTTTCACTGCTTACCCCTACGCGAGCCGCTTTTCCCA 2580
2581 TATCTTAAGGTAATTTTGGAAATCCCAATTTTAAACGATTAAGACCGTACGAGCTTCC 2640
2581 TATCTTAAGGTAATTTTGGAAATCCCAATTTTAAACGATTAAGACCGTACGAGCTTCC 2640
2641 TGGGATTCGCTGGAGCATTTTACAAAATTTTACGCAATTTTACGCAATTTTACGCAATTTT 2700
2641 TGGGATTCGCTGGAGCATTTTACAAAATTTTACGCAATTTTACGCAATTTTACGCAATTTT 2700
2701 AACTCAACACTGATCAGATTAATTAATTTTCAATTAACCTTTTACGATGATTTGATGATCT 2760
2701 AACTCAACACTGATCAGATTAATTAATTTTCAATTAACCTTTTACGATGATTTGATGATCT 2760
2761 ATCTATGATCTTTTCTTACACGCGTGGATGAAAGTTATAGTATTAATTAATTAATTAAT 2820
2761 ATCTATGATCTTTTCTTACACGCGTGGATGAAAGTTATAGTATTAATTAATTAATTAAT 2820
2821 ACAATTTGATTAATGATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 2880
2821 ACAATTTGATTAATGATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 2880
2881 CATATTTGATCAGATCGAGCTTCTGCACTTTTGTATTAATTAATTAATTAATTAATTAAT 2940
2881 CATATTTGATCAGATCGAGCTTCTGCACTTTTGTATTAATTAATTAATTAATTAATTAAT 2940
2941 AAGTATTAAGGACATTAAGGACGCAACAAAGTATTAAGACGAGGAAACAAAGCA 3000
2941 AAGTATTAAGGACATTAAGGACGCAACAAAGTATTAAGACGAGGAAACAAAGCA 3000
3001 TGAACCTATGTTGTTAGTTAAGCTTAATTAAGAGATTTTAAATTAATTAATTAATTAAT 3060
3001 TGAACCTATGTTGTTAGTTAAGCTTAATTAAGAGATTTTAAATTAATTAATTAATTAAT 3060
3061 AATACATTAATTTTTCGATCTTTTAAACCCCTCTTACAAACGAAAGCTCCCTTTT 3120
3061 AATACATTAATTAATTTTTCGATCTTTTAAACCCCTCTTACAAACGAAAGCTCCCTTTT 3120
3121 TCAGTAAAGTCCGATCCCATCTTAAAGCAACAAAGCATTAAGAAAGAGAGAGAGAGAGAG 3180
3121 TCAGTAAAGTCCGATCCCATCTTAAAGCAACAAAGCATTAAGAAAGAGAGAGAGAGAGAG 3180
3181 GAG 3202
3181 GAG 3202

RESULT 2
US-09-502-426-1
Sequence 1, Application US/09502426
GENERAL INFORMATION:
APPLICANT: AZPIROZ, Ricardo
APPLICANT: CHOE, Sunghwa
APPLICANT: FELDMANN, Kenneth
TITLE OF INVENTION: DMF4 POLYNUCLEOTIDES, POLYPEPTIDES AND USES THEREOF
FILE REFERENCE: 2225-0001
CURRENT APPLICATION NUMBER: US/09/502,426
EARLIER FILING DATE: 2000-02-11
EARLIER APPLICATION NUMBER: 60/119,657
EARLIER FILING DATE: 1999-02-11
EARLIER APPLICATION NUMBER: 60/119,658
NUMBER OF SEQ ID NOS: 18
SOFTWARE: PatentIn Ver. 2.0
SEQ ID NO 1
LENGTH: 6888
TYPE: DNA
ORGANISM: Arabidopsis sp.

US-09-502-426-1

Query Match 100.0%; Score 3202; DB 19; Length: 6888;
 Best Local Similarity 100.0%; Pred. No. 0;
 Matches 3202; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 ATGTGGGATTAATATTTGTTGGTTCGTTGACGATCAATTAATTTGCTGTTCTGCT 60
 DB 1 ATGTGGGATTAATATTTGTTGGTTCGTTGACGATCAATTAATTTGCTGTTCTGCT 60
 QY 61 TATTCGTTACAGATTTGATTTGTTGCTCAATTTGATTCGCAAGATTAATTAAT 120
 DB 61 TATTCGTTACAGATTTGATTTGTTGCTCAATTTGATTCGCAAGATTAATTAAT 120
 QY 121 AAAATTCATTTAAATATTTTACAGATTAATTAATTTGATTCGATTTGTTATTA 180
 DB 121 AAAATTCATTTAAATATTTTACAGATTAATTAATTTGATTCGATTTGTTATTA 180
 QY 122 AAAATTCATTTAAATATTTTACAGATTAATTAATTTGATTCGATTTGTTATTA 180
 DB 122 AAAATTCATTTAAATATTTTACAGATTAATTAATTTGATTCGATTTGTTATTA 180
 QY 181 AATATCTATCTTTGGTATATGAGAAAATATGAGTTGGAATTTATTAATTAAGAAA 240
 DB 181 AATATCTATCTTTGGTATATGAGAAAATATGAGTTGGAATTTATTAATTAAGAAA 240
 QY 241 TAATCGATTCATTTGGTTGGATTTACACAGTTGTTGTTCTTTCTTTGTTATAT 300
 DB 241 TAATCGATTCATTTGGTTGGATTTACACAGTTGTTGTTCTTTCTTTGTTATAT 300
 QY 301 ATATGAGTAATCAAAAGATATGATTTGAATGTAACATATTTGTTGATGACCCCA 360
 DB 301 ATATGAGTAATCAAAAGATATGATTTGAATGTAACATATTTGTTGATGACCCCA 360
 QY 361 AAAAAAATTTGGAATTCATATCTGTTGATTAATTAATTAAGAGTGAATATCC 540
 DB 361 AAAAAAATTTGGAATTCATATCTGTTGATTAATTAATTAAGAGTGAATATCC 540
 QY 421 AGTTTATTTGATCATATATGATGATCATCTTTGATTTGATTTGATTTGATTT 480
 DB 421 AGTTTATTTGATCATATATGATGATCATCTTTGATTTGATTTGATTTGATTT 480
 QY 481 CAATTAATTTGGAATTCATATCTGTTGATTAATTAATTAAGAGTGAATATCC 540
 DB 481 CAATTAATTTGGAATTCATATCTGTTGATTAATTAATTAAGAGTGAATATCC 540
 QY 541 GTTTATCGATCATCAATCATATATATGATTTGTTGCTTAATTCGCAAAATTTATTA 600
 DB 541 GTTTATCGATCATCAATCATATATATGATTTGTTGCTTAATTCGCAAAATTTATTA 600
 QY 601 AGAGTATTTGAGAAAAACGAAAAATTAAGAAAAAGGAGTGAATGAGTGAATG 660
 DB 601 AGAGTATTTGAGAAAAACGAAAAATTAAGAAAAAGGAGTGAATGAGTGAATG 660
 QY 661 TGAATTAATTAATCAAGAAATTAAGATGACAAACAAAGTTGTGAATATGTCCT 720
 DB 661 TGAATTAATTAATCAAGAAATTAAGATGACAAACAAAGTTGTGAATATGTCCT 720
 QY 721 GCGAGTTTCTCACAATCAATATGACCTATTTGATTTCTGATATTTGTTAAA 780
 DB 721 GCGAGTTTCTCACAATCAATATGACCTATTTGATTTCTGATATTTGTTAAA 780
 QY 781 TTTGCAATACGATTTGAAAAATATTTTATTTGTTAGCTGATCTCAATATTAAT 840
 DB 781 TTTGCAATACGATTTGAAAAATATTTTATTTGTTAGCTGATCTCAATATTAAT 840
 QY 841 GGTATTTGCAATATCTTCTGTTTAAAGCAATTTTCTTTCTTTTCTTTCTCT 900
 DB 841 GGTATTTGCAATATCTTCTGTTTAAAGCAATTTTCTTTCTTTTCTTTCTCT 900
 QY 901 TAACTATATATTTGCGGATATATGATTAATGATATATCAAAACATTTGTCGG 960
 DB 901 TAACTATATATTTGCGGATATATGATTAATGATATATCAAAACATTTGTCGG 960
 QY 961 ACCATTTTGAATTAATCTTTTCTCAAAACATTAAGGACACTGAGACCTTAATTA 1020
 DB 961 ACCATTTTGAATTAATCTTTTCTCAAAACATTAAGGACACTGAGACCTTAATTA 1020

QY 1021 CGATTTTACAGCGCTACACTAGTTGAGATTACTAGCATTAAGCATTAAGACCGTTCAAGC 1080
 DB 1021 CGATTTTACAGCGCTACACTAGTTGAGATTACTAGCATTAAGCATTAAGACCGTTCAAGC 1080
 QY 1081 TATTTATACAAAGTTTACAAACCTGATATAGCTTGAATTCCTTTGAAAAATTTTGAATTA 1140
 DB 1081 TATTTATACAAAGTTTACAAACCTGATATAGCTTGAATTCCTTTGAAAAATTTTGAATTA 1140
 QY 1141 CCGGTTTATTTGTAATATATGATTTAGTGAATTAATTAATTAATTAATTAATTAATTA 1200
 DB 1141 CCGGTTTATTTGTAATATATGATTTAGTGAATTAATTAATTAATTAATTAATTAATTA 1200
 QY 1201 ACATATACATTAATTCCTTACAGAAAAACAACTTAAGAGATTTAATCAATATCAATATAT 1260
 DB 1201 ACATATACATTAATTCCTTACAGAAAAACAACTTAAGAGATTTAATCAATATCAATATAT 1260
 QY 1261 GGGTATGCTTATACCTTTACAGATGATGCTATATCTAGAGACTTAAGAAATGTTATGATGTC 1320
 DB 1261 GGGTATGCTTATACCTTTACAGATGATGCTATATCTAGAGACTTAAGAAATGTTATGATGTC 1320
 QY 1321 GATTAATGAATTCACAGCGGTGATTAATTAATTAATTAATTAATTAATTAATTAATTA 1380
 DB 1321 GATTAATGAATTCACAGCGGTGATTAATTAATTAATTAATTAATTAATTAATTAATTA 1380
 QY 1381 ATATCATTTCTTGGTTGCTCAACATTAATTAATTAATTAATTAATTAATTAATTAATTA 1440
 DB 1381 ATATCATTTCTTGGTTGCTCAACATTAATTAATTAATTAATTAATTAATTAATTAATTA 1440
 QY 1441 CTGATGATTCATCAATGATCTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 1500
 DB 1441 CTGATGATTCATCAATGATCTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 1500
 QY 1501 TCTACAAGCGTGAATACCATGTCACATTAATTAATTAATTAATTAATTAATTAATTAATTA 1560
 DB 1501 TCTACAAGCGTGAATACCATGTCACATTAATTAATTAATTAATTAATTAATTAATTAATTA 1560
 QY 1561 ATTTATTTGTTAGTTGAATTTAATTAAGAGCACTTTGTAACATTAATTAATTAATTA 1620
 DB 1561 ATTTATTTGTTAGTTGAATTTAATTAAGAGCACTTTGTAACATTAATTAATTAATTAATTA 1620
 QY 1621 ATATCATGATGATTAATTTCCAAATTAATTAATTAATTAATTAATTAATTAATTAATTA 1680
 DB 1621 ATATCATGATGATTAATTTCCAAATTAATTAATTAATTAATTAATTAATTAATTAATTA 1680
 QY 1681 TCCATAGGTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 1740
 DB 1681 TCCATAGGTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 1740
 QY 1741 ATGAAGATTAACCTTAATGACTTAATTTTGAATAATTAATTAATTAATTAATTAATTA 1800
 DB 1741 ATGAAGATTAACCTTAATGACTTAATTTTGAATAATTAATTAATTAATTAATTAATTA 1800
 QY 1801 AATTAACGATGATTTTGTGTCATATGACACCTCTTAACCTGATATGCTAATTTT 1860
 DB 1801 AATTAACGATGATTTTGTGTCATATGACACCTCTTAACCTGATATGCTAATTTT 1860
 QY 1861 TCTGCAAAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 1920
 DB 1861 TCTGCAAAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 1920
 QY 1921 ACAATTTAATTTAAGCAAAATTTTGAATAATTAATTAATTAATTAATTAATTAATTA 1980
 DB 1921 ACAATTTAATTTAAGCAAAATTTTGAATAATTAATTAATTAATTAATTAATTAATTAATTA 1980
 QY 1981 AATATGATGCTTATATGATTTTCTGATATGCTTAAATTAATTTTATTAATTAATTAATTA 2040
 DB 1981 AATATGATGCTTATATGATTTTCTGATATGCTTAAATTAATTTTATTAATTAATTAATTA 2040
 QY 2041 TAAATCATTAATTAACCAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 2100
 DB 2041 TAAATCATTAATTAACCAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 2100

2101 AATCTACAAATTTATTTATTTAGTCAATTAACAAATGATAGAAAGTTCACAAAATTT 2160
2101 AATCTACAAATTTATTTATTTAGTCAATTAACAAATGATAGAAAGTTCACAAAATTT 2160
2161 TTGTTAAGAAAGTTCACAAATTTTTTATTTATTTAGCAAGAAATTAACAGATGAAA 2220
2161 TTGTTAAGAAAGTTCACAAATTTTTTATTTATTTAGCAAGAAATTAACAGATGAAA 2220
2221 CTATTTTGTGTGAATGAGATGATATATATATATTAAGCAATTTTAAATTTATATA 2280
2221 CTATTTTGTGTGAATGAGATGATATATATATATTAAGCAATTTTAAATTTATATA 2280
2281 ACCCTATAGCGCTCAAAATGATATATATATATAGTGTATATATATATATATATAT 2340
2281 ACCCTATAGCGCTCAAAATGATATATATATATAGTGTATATATATATATATATAT 2340
2341 TCAGAAATGGGACACCAATGAAAGCAATTAATTAATTAATTAATTAATTAATTAAT 2400
2341 TCAGAAATGGGACACCAATGAAAGCAATTAATTAATTAATTAATTAATTAATTAAT 2400
2401 TTGAGTAAATGTTTCTGACTATTTAGGGGCAAAAAAGACAAATGCCAAAGTCTAC 2460
2401 TTGAGTAAATGTTTCTGACTATTTAGGGGCAAAAAAGACAAATGCCAAAGTCTAC 2460
2461 GGGTTTGTGCTGCGTCAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 2520
2461 GGGTTTGTGCTGCGTCAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 2520
2521 TAGGGGCTTCTGCAATTTTCACTGTTCTACCCCTCTGCTGAGCCGACCCCTTTCCA 2580
2521 TAGGGGCTTCTGCAATTTTCACTGTTCTACCCCTCTGCTGAGCCGACCCCTTTCCA 2580
2581 TATCTAAGGTAATTTTGGAAATCCCAATTTAAACGATGAGACCGTACCGGACTTCC 2640
2581 TATCTAAGGTAATTTTGGAAATCCCAATTTAAACGATGAGACCGTACCGGACTTCC 2640
2641 TGGGATCTGCTGAGCAATTTATCAAAAATTTATAGACCAAGGGTTTATTAATTTAA 2700
2641 TGGGATCTGCTGAGCAATTTATCAAAAATTTATAGACCAAGGGTTTATTAATTTAA 2700
2701 AACCTACAACTGATCAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 2760
2701 AACCTACAACTGATCAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 2760
2761 ATCTAATGACTTTTCTTCTACACGAGTGAAGTTATAGTATTAAGCCAGAG 2820
2761 ATCTAATGACTTTTCTTCTACACGAGTGAAGTTATAGTATTAAGCCAGAG 2820
2821 ACAATGATATATGATATATCAATTAATTAATTAATTAATTAATTAATTAATTAAT 2880
2821 ACAATGATATATGATATATCAATTAATTAATTAATTAATTAATTAATTAATTAAT 2880
2881 CTATTTTGTGTGAATGAGATGATATATATATTAAGCAATTTTAAATTTATATA 2940
2881 CTATTTTGTGTGAATGAGATGATATATATATTAAGCAATTTTAAATTTATATA 2940
2941 AAGTATTAAGAGAGATTAACGAGCAAAAGTATGATGACGAGAAACAAAGCCA 3000
2941 AAGTATTAAGAGAGATTAACGAGCAAAAGTATGATGACGAGAAACAAAGCCA 3000
3001 TGAAGCTCATGTTAGTTAAGCTTAATTAAGAGATTTTATTAATTTATTAAGCATG 3060
3001 TGAAGCTCATGTTAGTTAAGCTTAATTAAGAGATTTTATTAATTTATTAAGCATG 3060
3061 ATTAACAATTAATTTTCTGACTTTTAAACCCCTCTTACAAACAGAGCTCCCTTTT 3120
3061 ATTAACAATTAATTTTCTGACTTTTAAACCCCTCTTACAAACAGAGCTCCCTTTT 3120
3121 TCAGTAAAGTCCGATTCATCTTAAAGCAAAAGCAATTAAGAAAGGATGATGCA 3180
3121 TCAGTAAAGTCCGATTCATCTTAAAGCAAAAGCAATTAAGAAAGGATGATGCA 3180
3181 GAGAGAGAGAGAACTAGCTCC 3202

Db 3181 GAGAGAGAGAGAACTAGCTCC 3202
RESULT 3
US-09-692-412-94/C
; Sequence 94, Application US/09692412
; GENERAL INFORMATION:
; APPLICANT: Bush, David F.
; APPLICANT: Rounsley, Steven D.
; APPLICANT: Wiegand, Roger C.
; TITLE OR INVENTION: Plant Polymorphic Markers and Uses Thereof
; FILE REFERENCE: 38-10(15493)C
; CURRENT APPLICATION NUMBER: US/09/692,412
; CURRENT FILING DATE: 2000-10-24
; NUMBER OF SEQ ID NOS: 124
; SEQ ID NO 94
; LENGTH: 84196
; TYPE: DNA
; ORGANISM: Arabidopsis thaliana
US-09-692-412-94

Query Match 97.9%; Score 3134.2; DB 27; Length 84196;
Best Local Similarity 99.5%; Pred. No. 0;
Matches 3186; Conservative 0; Mismatches 13; Indels 4; Gaps 4;

1 ATGTGGATTAATTAATTTGTTGGTTCGTTTGAACCTCAATTAATTAATTTGTTCTGTG 60
32199 ATGTGGATTAATTAATTTGTTGGTTCGTTTGAACCTCAATTAATTAATTTGTTCTGTG 32140
61 TATTCGTTCAATGATTTGAGTTGGTCTCAATTTGATTCGAATTAATTAATTAATTT 120
32139 TATTCGTTCAATGATTTGAGTTGGTCTCAATTTGATTCGAATTAATTAATTAATTT 32080
121 AAAATTCATTTAAATTAATTTTACAGTAATTAATTAATTTTACATTTGATTTGATTAACAA 180
32079 AAAATTCATTTAAATTAATTTTACAGTAATTAATTAATTTTACATTTGATTTGATTAACAA 32020
181 AATATCTACTTTGGTATATGCAAAATTAATTAATTAATTAATTAATTAATTAATTAAT 240
32019 AATATCTACTTTGGTATATGCAAAATTAATTAATTAATTAATTAATTAATTAATTAAT 31960
241 TATCGATTCATTTGGTGGTATGACAGTAACTTTTGGTCTTCTTTGTTATATGT 300
31959 TATCGATTCATTTGGTGGTATGACAGTAACTTTTGGTCTTCTTTGTTATATGT 31900
301 ATATGATTAATCAAAAGAGTATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 360
31899 ATATGATTAATCAAAAGAGTATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 31840
361 AAAAAAAG 420
31839 AAAAAAAG 31781
421 AGTTTATTTGATCAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 480
31780 AGTTTATTTGATCAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 31721
481 CAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 540
31720 CAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 31661
541 GTTATGATCACTCCATCAATCAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 600
31660 GTTATGATCACTCCATCAATCAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 31601
601 AGAGTATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 660
31600 AGAGTATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 31541
661 TGAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 720
31540 TGAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 31481

Db 30821 ATATCATCTTGTGGTCAACAATTAACAAAAACAGAAAAACGATTTT 30762
Qy 1441 CTGGATTCATTCATGATCTAAATGATAGATCTTTGGTTAGAGTTCCGAGTCC 1500
Db 30761 CTGGATTCATTCATGATCTAAATGATAGATCTTTGGTTAGAGTTCCGAGTCC 30702
Qy 1501 TCTCAAGCGGTACCATCTGCAACTATTAAATGCTTTCTTTAAATGATCTTTAAACAT 1560
Db 30701 TCTCAAGCGGTACCATCTGCAACTATTAAATGCTTTCTTTAAATGATCTTTAAACAT 30642
Qy 1561 ATTATTTGTAGTGAATTAATAGAGCACTGTGAACATTACATATTATTTAG 1620
Db 30641 ATTATTTGTAGTGAATTAATAGAGCACTGTGAACATTACATATTATTTAG 30582
Qy 1621 ATACTAGTATGTATATTCCAATACACTTGTGAATGTTAACTTAATCTTGTCT 1680
Db 30581 ATACTAGTATGTATATTCCAATACACTTGTGAATGTTAACTTAATCTTGTCT 30522
Qy 1681 TCCAGCGTATTAATTAATGATGAGTAAAAAGTTTGTCTATTTCGCGATGC 1740
Db 30521 TCCAGCGTATTAATTAATGATGAGTAAAAAGTTTGTCTATTTCGCGATGC 30463
Qy 1741 ATGAGGATTAACCTTAATGATGATTTTGAAGATGTAACCTTTTACTCATAGAT 1800
Db 30462 ATGAGGATTAACCTTAATGATGATTTTGAAGATGTAACCTTTTACTCATAGAT 30403
Qy 1801 AATTACCGTATGTTTGTGCTCAATAGACGCTCTACACCTGTGATGTAATTTT 1860
Db 30402 AATTACCGTATGTTTGTGCTCAATAGACGCTCTACACCTGTGATGTAATTTT 30343
Qy 1861 TCTGCAATTAATTAATGAGTAATGATGATGATGATGATGATGATGATGAT 1920
Db 30342 TCTGCAATTAATTAATGAGTAATGATGATGATGATGATGATGATGATGAT 30283
Qy 1921 ACATTTTAATTTAAAGCAAAATTTTGAAGATGTAATTAATTTCTAACAATATTATTA 1980
Db 30282 ACATTTTAATTTAAAGCAAAATTTTGAAGATGTAATTAATTTCTAACAATATTATTA 30223
Qy 1981 AATATGATGCTATATGATGATGATGATGATGATGATGATGATGATGATGAT 2040
Db 30222 AATATGATGCTATATGATGATGATGATGATGATGATGATGATGATGATGAT 30163
Qy 2041 TAATATCATTAATGATGATGATGATGATGATGATGATGATGATGATGATGAT 2100
Db 30162 TAATATCATTAATGATGATGATGATGATGATGATGATGATGATGATGATGAT 30103
Qy 2101 AATATCAAAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 2160
Db 30102 AATATCAAAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 30043
Qy 2161 TTGTTAAGCAAACTTCCAAATTTTTTTTTTATGAGACAGAAATTAAGATAGAAA 2220
Db 30042 TTGTTAAGCAAACTTCCAAATTTTTTTTTTATGAGACAGAAATTAAGATAGAAA 22983
Qy 2221 CTATTTGTGTGGAATGATGATGATGATGATGATGATGATGATGATGATGAT 2280
Db 22982 CTATTTGTGTGGAATGATGATGATGATGATGATGATGATGATGATGATGAT 22923
Qy 2281 AGCCTATACGCGCTCAAGATGATGATGATGATGATGATGATGATGATGATGAT 2340
Db 22922 AGCCTATACGCGCTCAAGATGATGATGATGATGATGATGATGATGATGATGAT 22963
Qy 2341 TCGAATTTGGGACAAACATGAAAAAGGAAATTAATTAATTAATTAATTAAT 2400
Db 22962 TCGAATTTGGGACAAACATGAAAAAGGAAATTAATTAATTAATTAATTAAT 22903
Qy 2401 TTGAGTAAATGTGTTTCTGATGATGATGATGATGATGATGATGATGATGAT 2460
Db 22902 TTGAGTAAATGTGTTTCTGATGATGATGATGATGATGATGATGATGATGAT 22943
Qy 2461 GGGTTTGAATGCTGATGATGATGATGATGATGATGATGATGATGATGATGAT 2520
Db 22942 GGGTTTGAATGCTGATGATGATGATGATGATGATGATGATGATGATGATGAT 22983

Qy 2521 TAGGGTCTCTGACATTTTCACTGTTCTACCCCTACTCGTGAGCCACCCTTTTCCCA 2580
Db 22982 TAGGGTCTCTGACATTTTCACTGTTCTACCCCTACTCGTGAGCCACCCTTTTCCCA 22963
Qy 2581 TATCTTAAGGTAATTTTGGAAATCCCAATTTAAACGATTTGAGACCTTACCGGACTTCC 2640
Db 22962 TATCTTAAGGTAATTTTGGAAATCCCAATTTAAACGATTTGAGACCTTACCGGACTTCC 22963
Qy 2641 TGGGATTCCTGCTGGAGCATTTATCAAAAATTAATTAATTAATTAATTAATTAAT 2700
Db 22962 TGGGATTCCTGCTGGAGCATTTATCAAAAATTAATTAATTAATTAATTAATTAAT 22953
Qy 2701 AACTCACAATCTGATCAGATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 2760
Db 22952 AACTCACAATCTGATCAGATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 22943
Qy 2761 ATCTAATGATCTTTTCTTCTACAGCGGTGATGATGATGATGATGATGATGATGAT 2820
Db 22942 ATCTAATGATCTTTTCTTCTACAGCGGTGATGATGATGATGATGATGATGATGAT 229383
Qy 2821 ACAATGATTAATGATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 2880
Db 229382 ACAATGATTAATGATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 229323
Qy 2881 CTATTTGACATGCGACCTTCTGCAACTTTTGTGTTTAAATTAATTAATTAATTAAT 2940
Db 229322 CTATTTGACATGCGACCTTCTGCAACTTTTGTGTTTAAATTAATTAATTAATTAAT 229263
Qy 2941 AAGTATTAAGGAGCAATTAAGGAGCAAAACCAATTAATTAATTAATTAATTAATTAAT 3000
Db 229262 AAGTATTAAGGAGCAATTAAGGAGCAAAACCAATTAATTAATTAATTAATTAATTAAT 229203
Qy 3001 TGAAGCTCATGTTGTTA-GTTTAAAGCTTAATTAATTAATTAATTAATTAATTAATTAAT 3059
Db 229202 TGAAGCTCATGTTGTTAAGCTTAATTAATTAATTAATTAATTAATTAATTAATTAAT 229143
Qy 3060 GATTAACAATTAATTTCTGACTTTTAAACCCCTCTTAAACAGAAAGTCCCTTT 3119
Db 229142 GATTAACAATTAATTTCTGACTTTTAAACCCCTCTTAAACAGAAAGTCCCTTT 229083
Qy 3120 TTGAGTGAAGTCCGATTTCCCAATTAATTAATTAATTAATTAATTAATTAATTAAT 3179
Db 229082 TTGAGTGAAGTCCGATTTCCCAATTAATTAATTAATTAATTAATTAATTAATTAAT 229023
Qy 3180 AGAGAGAGAGAGAACTAGCTCC 3202
Db 229022 AGAGAGAGAGAGAACTAGCTCC 229000

RESULT 5
US-09-451-320-2050
Sequence 2050, Application US/09451320
GENERAL INFORMATION:
APPLICANT: ALEXANDROV, Nikolai
APPLICANT: CHEN, Xianfeng
APPLICANT: SUBRAMANIAN, Gopalakrishnan
APPLICANT: ZHENG, Liansheng
TITLE OF INVENTION: SEQUENCE-DETERMINED DNA FRAGMENTS 42-59, 62-66, 68, 81
FILE REFERENCE: 2750-0662P
CURRENT APPLICATION NUMBER: US/09/451,320
NUMBER OF SEQ ID NOS: 6998
SOFTWARE: PatentIn Ver. 2.1
SEQ ID NO 2050
LENGTH: 1699
TYPE: DNA
ORGANISM: Arabidopsis thaliana
FEATURE:
OTHER INFORMATION: any n or xaa - unknown
US-09-451-320-2050
Query Match 30.0%; Score 960.4; DB 18; Length 1699;

QY 942 CACAAACAA 951
DB 33340 TTACAAAAA 33349

RESULT 8

US-09-803-736-1074
; Sequence 1074, Application US/09803736
; GENERAL INFORMATION:
; APPLICANT: Bush, David F.
; APPLICANT: Levin, Irena M.
; APPLICANT: Norris, Susan R.
; APPLICANT: Rounsley, Steven D.
; APPLICANT: Wiegand, Roger C.
; TITLE OF INVENTION: Plant Polymorphic Markers and Uses Thereof
; FILE REFERENCE: 38-10(15493)D
; CURRENT APPLICATION NUMBER: US/09/803/736
; CURRENT FILING DATE: 2001-03-12
; PRIOR APPLICATION NUMBER: US 09/534,859
; PRIOR FILING DATE: 2000-03-29
; PRIOR APPLICATION NUMBER: Identified by Attorney Docket number 04983.0206CPUS01 38-10
; PRIOR FILING DATE: 2000-10-20
; NUMBER OF SEQ ID NOS: 1582
; SEQ ID NO 1074
; LENGTH: 84196
; TYPE: DNA
; ORGANISM: Arabidopsis thaliana
US-09-803-736-1074

Query Match 3.7%; Score 118; DB 31; Length 84196;
Best Local Similarity 83.2%; Pred. No. 3.7e-07;
Matches 158; Conservative 0; Mismatches 30; Indels 2; Gaps 2;

QY 763 TCTGATATTCGTTAAATTTGGATACGATGTGAAATATTTATTTGTTAGCTGA 822
DB 33161 TATAGAAATTCGTTAAATTTGGATACGATGTGAAATATTTATTTGTTAGCTGA 33220
QY 823 TCTCAATATTTATGTTCCAGGATTTTGGATATCTCTGTTAAAGCATA-TTTTGTCTT 881
DB 33221 TCTCAATATTTATGTTCCAGGATTTTGGATATCTCTGTTAAAGCATA-TTTTGTCTT 33279
QY 882 CTTTTCGTTCTGTTCTTACATATTTATTCGCGATATATGATTAACAATGATATAT 941
DB 33280 CTTTTCGTTCTGTTCTTACATATTTATTCGCGATATATGATTAACAATGATATAT 33339
QY 942 CACAAACAA 951
DB 33340 TTACAAAAA 33349

RESULT 9

US-60-207-458-33653
; Sequence 33653, Application US/60207458
; GENERAL INFORMATION:
; APPLICANT: Abad, Mark S.
; APPLICANT: Conner, Timothy W.
; APPLICANT: Delkman, Jill
; APPLICANT: Hardeman, Kristine J.
; APPLICANT: La Rosa, Thomas J.
; APPLICANT: Lalugudi, Raghunath V.
; APPLICANT: Ruan, Yijun G.
; APPLICANT: Ruff, Thomas G.
; APPLICANT: Sammons, R. Douglas
; APPLICANT: Shukla, Hridayabhiranjan
; APPLICANT: Wu, Kunsheng
; APPLICANT: Xu, Nanfei
; TITLE OF INVENTION: NUCLEIC ACID MOLECULES AND OTHER MOLECULES ASSOCIATED WITH
; FILE REFERENCE: 38-21(51936)A
; CURRENT APPLICATION NUMBER: US/60/207,458
; CURRENT FILING DATE: 2000-05-30
; NUMBER OF SEQ ID NOS: 152403
; SEQ ID NO 33653

LENGTH: 1016
; TYPE: DNA
; ORGANISM: Zea mays
; OTHER INFORMATION: Clone ID: uc-zmflb73263c09b1
US-60-207-458-33653

Query Match 3.3%; Score 105.6; DB 64; Length 1016;
Best Local Similarity 49.3%; Pred. No. 1.3e-05;
Matches 452; Conservative 0; Mismatches 444; Indels 20; Gaps 6;

QY 1527 TATTAATTCGTTCTTAAATGCAATCTTAAACATATTTATTTGTTAGTGAATTAATA 1586
DB 79 TATTAATTCGTTCTTAAATGCAATCTTAAACATATTTATTTGTTAGTGAATTAATA 138
QY 1587 GAGCGAATCTGTAACATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 1646
DB 139 TTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 198
QY 1647 CATACCTTGGATGTTAAACCTTAATCTGTTCTTCCAGCTATTAATA-TTAATCAT 1764
DB 199 TTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 258
QY 1705 CGAGCTAAATTAATTTGCTTATTTTCGGAATGAAGATTAACCTTAATGACTT 1764
DB 259 TTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 318
QY 1765 AATTTTGAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 1824
DB 319 TATTTTAAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 377
QY 1825 TAAATGACGCTC--TACACCTGTAATGTAATTTTTCGCAATTAATTAATTAATGA 1881
DB 378 TATTTTAAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 437
QY 1882 ATTCATGCTACTTCATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 1941
DB 438 AATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 492
QY 1942 ATTTTGAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 1993
DB 493 ATATTTTAAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 552
QY 1994 TAAATGATTCCTTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 2053
DB 553 TATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 612
QY 2054 AACCAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 2113
DB 613 ATATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 672
QY 2114 ATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 2173
DB 673 ATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 732
QY 2174 CTTCCAAATTTTATTTTATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 2233
DB 733 ATATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 792
QY 2234 GAATGAAGTAGTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 2292
DB 793 ATATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 852
QY 2293 CTCGAAGTAGTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 2352
DB 853 TTTATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 912
QY 2353 CAACAAATGAAGCAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 2412
DB 913 TTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 972
QY 2413 GTTTTCGACTATTTGA 2428
DB 973 TTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 988

```

RESULT 10
US-09-663-779-8205/C
; Sequence 8205, Application US/09663779
; GENERAL INFORMATION:
; APPLICANT: Corbin, David R.
; APPLICANT: Malvar, Thomas M.
; APPLICANT: Shukla, Hridayabhijan
; TITLE OF INVENTION: BACILLUS THURINGIENSIS CHROMOSOMAL GENOME SEQUENCES AND USE
; TITLE OF INVENTION: THEREOF
; FILE REFERENCE: 38-21(51376)B
; CURRENT APPLICATION NUMBER: US/09/663,779
; CURRENT FILING DATE: 2000-09-15
; PRIOR APPLICATION NUMBER: 60/154,678
; PRIOR FILING DATE: 1999-09-17
; NUMBER OF SEQ ID NOS: 8283
; SEQ ID NO 8205
; LENGTH: 1326
; TYPE: DNA
; ORGANISM: Bacillus thuringiensis
; FEATURE:
; NAME/KEY: unsure
; LOCATION: (1)...(1326)
; OTHER INFORMATION: unsure at all n locations
US-09-663-779-8205

```

```

Query Match      3.2%; Score 102.4; DB 26; Length 1326;
Best Local Similarity 46.1%; Pred. No. 3.8e-05;
Matches 561; Conservative 0; Mismatches 642; Indels 14; Gaps 6;

```

```

QY 1211 AATCCCTTACAGAAAAAACAACCTTAGAGAGATTACATATCCATATATGGTATGCTA 1270
DB 1318 AAAATATATATATATATATATATATATATATATATATATATATATATATATAT 1259
QY 1271 TACCTTCAGTATCTTACTAGAGACTAAAGATCTTATGATGCTGATTAATGAA 1330
DB 1258 TAAATTAATATATATATATATATATATATATATATATATATATATATATATAT 1199
QY 1331 ATTCACACGCGTGTATATATATATATATATATATATATATATATATATATATAT 1390
DB 1198 ATATATATATATATATATATATATATATATATATATATATATATATATATATAT 1139
QY 1391 TGGTTGGTCAACAATTAACCAACCAACCAACCAACCAACCAACCAACCAACCAATTC 1450
DB 1138 TTTATATATATATATATATATATATATATATATATATATATATATATATATATAT 1081
QY 1451 ATTCAGATGCTAAATGCAATGCTTGGTGTACGTTGCAAGTCCCTGACAGCG 1510
DB 1080 ATTATATATATATATATATATATATATATATATATATATATATATATATATATAT 1021
QY 1511 TGTAAACATCTGCAACATTAATTAATGCTTCTTAATGCACTTAAACATATATATGTT 1570
DB 1020 TATATATATATATATATATATATATATATATATATATATATATATATATATATAT 961
QY 1571 AGTTGGAATTAATAAGGCAACTGTTGACATTAACATATATATATATATAGTACTAGT 1630
DB 960 ATTATATATATATATATATATATATATATATATATATATATATATATATATATAT 901
QY 1631 GTGATTAATCCAAATFACCTTGTGATGTTAACTTAACCTGTTCTCCACGCTA 1690
DB 900 TTATATATATATATATATATATATATATATATATATATATATATATATATATAT 841
QY 1691 TAAATATTAATCATGAGTAAAAAAGTTTGTCTTAATTTCCGATGCAAGTGAAGATA 1750
DB 840 ATATATATATATATATATATATATATATATATATATATATATATATATATATAT 781
QY 1751 AACCTAATGACTTATATATATATATATATATATATATATATATATATATATATAT 1810
DB 780 TATATATATATATATATATATATATATATATATATATATATATATATATATATAT 721
QY 1811 TGTATTTGTTGCCATTAATGACAGGCTTACAACTGTGATGTCAATTTTTCGCCAATA 1870

```

```

DB 720 AATATATATATATATATATATATATATATATATATATATATATATATATATAT 661
QY 1871 TTAATATGAGATTCATAGCTACATCAATAGACAAACAGCTGATATATATATATAT 1930
DB 660 TAAATATATATATATATATATATATATATATATATATATATATATATATATATAT 602
QY 1931 TTAAGACAAATTTTGGAAAAATCTTATATATATATATATATATATATATATATAT 1990
DB 601 TTAATATATATATATATATATATATATATATATATATATATATATATATATATAT 542
QY 1991 CTATATATATATATATATATATATATATATATATATATATATATATATATATAT 2050
DB 541 TATATATATATATATATATATATATATATATATATATATATATATATATATATAT 482
QY 2051 ATGAACCAATATATATATATATATATATATATATATATATATATATATATATAT 2110
DB 481 TTTATATCTTAAATATATATATATATATATATATATATATATATATATATATAT 422
QY 2111 ATTATATATATATATATATATATATATATATATATATATATATATATATATATAT 2170
DB 421 TATATATATATATATATATATATATATATATATATATATATATATATATATATAT 369
QY 2171 AACTTCCAAATTTTATATATATATATATATATATATATATATATATATATATATAT 2230
DB 368 ATATATATATATATATATATATATATATATATATATATATATATATATATATATAT 309
QY 2231 GTGAAATGCAAGTATATATATATATATATATATATATATATATATATATATATAT 2290
DB 308 TAAATATATATATATATATATATATATATATATATATATATATATATATATATAT 249
QY 2291 CGCTCA-AAGATATATATATATATATATATATATATATATATATATATATATAT 2349
DB 248 TTTATATATATATATATATATATATATATATATATATATATATATATATATATAT 191
QY 2350 GGACACAAATGCAAAAGCAATTAATATATATATATATATATATATATATATATATAT 2409
DB 190 AAATATATATATATATATATATATATATATATATATATATATATATATATATATAT 132
QY 2410 TGTGTTTCTGACTAT 2426
DB 131 ATTATATATATATATATATATATATATATATATATATATATATATATATATATAT 115

```

```

RESULT 11
US-09-663-779-8205
; Sequence 8205, Application US/09663779
; GENERAL INFORMATION:
; APPLICANT: Corbin, David R.
; APPLICANT: Malvar, Thomas M.
; APPLICANT: Shukla, Hridayabhijan
; TITLE OF INVENTION: BACILLUS THURINGIENSIS CHROMOSOMAL GENOME SEQUENCES AND
; TITLE OF INVENTION: THEREOF
; FILE REFERENCE: 38-21(51376)B
; CURRENT APPLICATION NUMBER: US/09/663,779
; CURRENT FILING DATE: 2000-09-15
; PRIOR APPLICATION NUMBER: 60/154,678
; PRIOR FILING DATE: 1999-09-17
; NUMBER OF SEQ ID NOS: 8283
; SEQ ID NO 8205
; LENGTH: 1326
; TYPE: DNA
; ORGANISM: Bacillus thuringiensis
; FEATURE:
; NAME/KEY: unsure
; LOCATION: (1)...(1326)
; OTHER INFORMATION: unsure at all n locations
US-09-663-779-8205

```

```

Query Match      3.1%; Score 98.2; DB 26; Length 1326;
Best Local Similarity 45.6%; Pred. No. 0.00015;
Matches 426; Conservative 0; Mismatches 503; Indels 6; Gaps 2;

```

```

1528 ATTAATTTGCTTCTTAATGCACTTTAAACATATATATATATATATATATATATATATAG 1587

```


Db

593 TTTTAAATATTTAAATTTATATATATTTAAATATATATAATATATA 534

1658 TGTGTTAAACCTTAATCTGTGTTCTTCTTCTACGGTATTAATATATATCAT ---CGAGGTAAA 1713


```

QY 1714 AAAAGTTTGCTTATTTTGGCGATGAGAGATTAACCTAATGACTTTAATTTTGG 1773
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 533 TATTAATATATATATATATATATATATATATATATATATATATATATATAT 474
QY 1774 AAAATGTAACCTTACATGATGATTAATTAATTAATTAATTAATTAATTAAT 1833
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 473 TATTAATATATATATATATATATATATATATATATATATATATATATAT 414
QY 1834 CCTCTCAACGTGATGATGATGATGATGATGATGATGATGATGATGATGAT 1893
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 413 ATATTAATATATATATATATATATATATATATATATATATATATATATAT 354
QY 1894 TATCAATGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1953
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 353 TATTAATATATATATATATATATATATATATATATATATATATATATAT 294
QY 1954 TGTATAATTTTCAACAAATATATATATATATATATATATATATATATAT 2013
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 293 ATTAATATATATATATATATATATATATATATATATATATATATATATAT 234
QY 2014 TAAATATATATATATATATATATATATATATATATATATATATATATAT 2073
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 233 TATTAATATATATATATATATATATATATATATATATATATATATATAT 174
QY 2074 TTTCAATATATATATATATATATATATATATATATATATATATATATAT 2133
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 173 TTAATATATATATATATATATATATATATATATATATATATATATATAT 114
QY 2134 AATGATATAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2193
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 113 AATATATATATATATATATATATATATATATATATATATATATATATAT 55
QY 2194 ATGGAACAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2208
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 54 ATTTATATATATATATATATATATATATATATATATATATATATATATAT 40

```

```

; FEATURE:
; NAME/KEY: misc.feature
; LOCATION: 807, 808, 819, 820, 821, 830, 838, 840, 844, 845, 849, 861,
; LOCATION: 871, 874, 875, 891, 892, 895, 898, 900, 901, 902, 904, 907,
; LOCATION: 917, 923, 925, 926, 927, 929, 955
; OTHER INFORMATION: n = A,T,C or G
US-10-198-846-6381

Query Match      2.9%; Score 91.6; DB 41; Length 960;
Best Local Similarity 39.5%; Pred. No. 0.0013;
Matches 357; Conservative 0; Mismatches 539; Indels 7; Gaps 1;

QY 1413 AAAACAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1472
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 959 ATATNAAAATATATATATATATATATATATATATATATATATATATATAT 900
QY 1473 GATCTTTGGGTACAGTTTCCAGAGTCTCTACAGGCTGTACATCTGCACTATA 1532
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 899 AATTTNAAAATATATATATATATATATATATATATATATATATATATATAT 840
QY 1533 ATTCCTTTCTTATGATGATGATGATGATGATGATGATGATGATGATGATG 1592
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 839 TNAATTTTNTATATATATATATATATATATATATATATATATATATATATAT 780
QY 1593 ACTTGATACATTAATATATATATATATATATATATATATATATATATATAT 1652
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 779 TTTTAAAAAATATATATATATATATATATATATATATATATATATATATAT 720
QY 1653 TTGAGATGTTAACTTAATCTTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 1712
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 719 TTTTAAAAAATATATATATATATATATATATATATATATATATATATATAT 660
QY 1713 AAAAGTTTGTCTTATTTTGGAGATGAGAGATTAACCTAATGACTTTAATTTT 1772
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 659 NNANTTTTTTTNTNANANAT-----TTAAAAAANNTTTTTTTTTTATANTTT 607
QY 1773 GAAAATGTAAACCTTTTACTCATATATATATATATATATATATATATATAT 1832
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 606 ATAAAAAATTTTNTTNTTNTTNTTNTTNTTNTTNTTNTTNTTNTTNTTNTT 547
QY 1833 GCCTGTACACCTGTATGATGATGATGATGATGATGATGATGATGATGATGAT 1892
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 546 AAACCTTNAATTAANANANATTTTNTTNTTNTTNTTNTTNTTNTTNTTNTT 487
QY 1893 CTATCATATGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1952
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 486 TTTATATATATATATATATATATATATATATATATATATATATATATATAT 427
QY 1953 ATGTTATATTTCTAACATATATATATATATATATATATATATATATATATAT 2012
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 426 AANANAAATTTTAAATNTNANANAAATTTTAAATNTNANANAAATTTTAAAC 367
QY 2013 TTAATAATTTTNTTNTTNTTNTTNTTNTTNTTNTTNTTNTTNTTNTTNTT 2072
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 366 NTCATTTNTTNTTNTTNTTNTTNTTNTTNTTNTTNTTNTTNTTNTTNTTNTT 307
QY 2073 ATTCATATATCTCATATATATATATATATATATATATATATATATATATATAT 2132
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 306 TTAATAAANNTAATAATNTATATAAGNNNTNTATNTTAAATAATATNTANNA 247
QY 2133 CAATGATAGAAAGTTCCAAAAATTTTGTATACAGAAACTCCAAATTTTTTTTT 2192
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 246 AATTTATATTTTATTTTNNAAAAATTTTAAAAAATTTTATTTTAAACNATGTGA 187
QY 2193 TATGAGAACAGAAATATAGATGAGAAACATTTTGTGGAATGGAAGTACTAATATA 2252
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 186 TAAAAAANNTTAAANANATAGNATTTTGAATNTTAAATNTTAAATATCTTTTAAATTA 127
QY 2253 CATTAAGCAAAATTTTAAAAAATATATATATATATATATATATATATATATAT 2312
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 126 ANTATATNTNTTTTAAAAAATTTTAAAAAATTTTAAAAAATTTTAAAAAATTTT 67
QY 2313 AGG 2315

```


Wed Jun 11 10:32:16 2003

us-09-502-426b-1_copy_1_3202.rmpm

Page 15

Db 66 NGG 64

Search completed: June 11, 2003, 05:00:41
Job time : 7696.77 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: June 10, 2003, 22:22:54 ; Search time 2372.13 Seconds
(without alignments)
8960.952 Million cell updates/sec

Title: US-09-502-426B-1_COPY_1_3202

Perfect score: 3202
Sequence: 1 atgtgggtatattatgtgtg.....gagagagagaactagctcc 3202

Scoring table: IDENTITY_MUC
Gapop 10.0 , Gapext 1.0

Searched: 7602234 seqs, 3319262570 residues

Total number of hits satisfying chosen parameters: 15204468

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Listing first 45 summaries

Database :

Pending Patents: NA_New: *
1: /cgn2_6/ptodata/1/pna/US01_NEW_COMB.seq: *
2: /cgn2_6/ptodata/1/pna/US07_NEW_COMB.seq: *
3: /cgn2_6/ptodata/1/pna/US08_NEW_COMB.seq: *
4: /cgn2_6/ptodata/1/pna/US09_NEW_COMB.seq: *
5: /cgn2_6/ptodata/1/pna/US09_NEW_COMB.seq: *
6: /cgn2_6/ptodata/1/pna/US09_NEW_COMB.seq: *
7: /cgn2_6/ptodata/1/pna/US09_NEW_COMB.seq: *
8: /cgn2_6/ptodata/1/pna/US10_NEW_COMB.seq: *
9: /cgn2_6/ptodata/1/pna/US10_NEW_COMB.seq: *
10: /cgn2_6/ptodata/1/pna/US60_NEW_COMB.seq: *
11: /cgn2_6/ptodata/1/pna/US60_NEW_COMB.seq: *

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	3202	100.0	6888	5	US-09-502-426B-1
2	3202	100.0	6888	7	US-09-502-426A-1
3	113.4	3.5	2111522	7	US-09-947-911-63
4	103	3.2	3673778	9	US-10-312-841-1
5	100.2	3.1	3673778	9	US-10-312-841-1
6	99.4	3.1	318007	6	US-09-948-124-61
7	99.4	3.1	600322	7	US-09-947-911-2
8	97.4	3.0	2453797	7	US-09-947-911-331
9	97	3.0	23939	10	US-60-466-412-84116
10	96	3.0	79363	10	US-60-466-412-84408
11	94.8	3.0	318007	6	US-09-948-124-61
12	94.8	3.0	600322	7	US-09-947-911-2
13	94.2	2.9	597573	7	US-09-947-911-179
14	92.4	2.9	4063095	7	US-09-947-911-323
15	91.6	2.9	1202549	10	US-60-465-241-51800
16	91.4	2.9	1599662	7	US-09-947-911-108
17	90.8	2.8	597573	7	US-09-947-911-179
18	90.6	2.8	7928029	7	US-09-947-916-16
19	90	2.8	24321	10	US-60-466-412-84035
20	90	2.8	31630	10	US-60-466-412-85572
21	90	2.8	4604723	7	US-09-947-916-240
22	89.2	2.8	656902	7	US-09-947-911-231

23	88	2.7	2682138	7	US-09-947-911-255	Sequence 255, App
24	87.8	2.7	4384464	6	US-09-948-124-7	Sequence 7, Appl
25	87.8	2.7	5490492	7	US-09-947-911-237	Sequence 237, App
26	87.2	2.7	511017	10	US-60-466-412-84564	Sequence 84564, A
27	86.6	2.7	656902	7	US-09-947-911-231	Sequence 231, App
28	86.2	2.7	49298	10	US-60-466-412-85631	Sequence 85631, A
29	86.2	2.7	3037227	7	US-09-947-911-286	Sequence 286, App
30	86.2	2.7	8616041	7	US-09-947-916-174	Sequence 174, App
31	85.6	2.7	13859	5	US-09-950-0838-10355	Sequence 10355, A
32	85.6	2.7	68406	10	US-60-466-412-86557	Sequence 86557, A
33	85.6	2.7	366701	7	US-09-947-911-254	Sequence 254, App
34	85	2.7	1391915	10	US-60-466-412-87958	Sequence 87958, A
35	84.6	2.6	9810	7	US-10-311-455-400	Sequence 400, App
36	84	2.6	18773	7	US-09-949-016-14164	Sequence 14164, A
37	84	2.6	26772	10	US-60-465-241-52473	Sequence 52473, A
38	84	2.6	26772	10	US-60-466-412-86156	Sequence 86156, A
39	84	2.6	54642	10	US-60-465-241-52561	Sequence 52561, A
40	84	2.6	54642	10	US-60-466-412-86282	Sequence 86282, A
41	84	2.6	1391915	10	US-60-466-412-87958	Sequence 87958, A
42	84	2.6	2347870	7	US-09-947-911-294	Sequence 294, App
43	83.8	2.6	8170	9	US-10-240-453-131	Sequence 131, App
44	83.8	2.6	1108228	10	US-60-466-412-86602	Sequence 86602, A
45	83.4	2.6	731264	7	US-09-947-911-156	Sequence 156, App

ALIGNMENTS

RESULT 1	US-09-502-426B-1
Sequence 1, Application US/09502426B	
GENERAL INFORMATION:	
APPLICANT: Applioz, Ricardo	
APPLICANT: Choe, Sunghwa	
APPLICANT: Feldman, Kenneth A.	
TITLE OF INVENTION: DMF4 POLYNUCLEOTIDES, POLYPEPTIDES AND USES THEREOF	
FILE REFERENCE: 11696-070001	
CURRENT APPLICATION NUMBER: US/09/502,426B	
CURRENT FILING DATE: 2000-02-11	
PRIOR APPLICATION NUMBER: US 60/119,657	
PRIOR FILING DATE: 1999-02-11	
PRIOR APPLICATION NUMBER: US 60/119,658	
PRIOR FILING DATE: 1999-02-11	
NUMBER OF SEQ ID NOS: 30	
SOFTWARE: FastSeq for Windows Version 4.0	
SEQ ID NO 1	
LENGTH: 6888	
TYPE: DNA	
ORGANISM: Arabidopsis thaliana	
US-09-502-426B-1	
Query Match	100.0%; Score 3202; DB 5; Length 6888;
Best Local Similarity	100.0%; Pred. No. 0;
Matches 3202; Conservative	0; Mismatches 0; Indels 0; Gaps 0;
1	ATGTGGTATATATGTTGGGTTGGTTGGACCTACAAATATAATTTCTGTTCTGCT 60
1	ATGTGGTATATATATGTTGGGTTGGTTGGACCTACAAATATAATTTCTGTTCTGCT 60
61	TATTCGTTACATGATTTGAGTTGCTCTCAATTTGATTCAGATTAATTAATTT 120
61	TATTCGTTACATGATTTGAGTTGCTCTCAATTTGATTCAGATTAATTAATTT 120
61	TATTCGTTACATGATTTGAGTTGCTCTCAATTTGATTCAGATTAATTAATTT 120
121	AAATATCATTTAAATATTTACAGTAATTAATTTATCTTATTTGTTATTAACA 180
121	AAATATCATTTAAATATTTACAGTAATTAATTTATCTTATTTGTTATTAACA 180
121	AAATATCATTTAAATATTTACAGTAATTAATTTATCTTATTTGTTATTAACA 180
181	AAATATCATTTGATTTGATTTGAGTAATTAATTTGAGTTGATTTGATTTGATTT 240
181	AAATATCATTTGATTTGATTTGAGTAATTAATTTGAGTTGATTTGATTTGATTT 240
181	AAATATCATTTGATTTGATTTGAGTAATTAATTTGAGTTGATTTGATTTGATTT 240
241	TATTCATTCATTTGATTTGATTTGATTTGATTTGATTTGATTTGATTTGATTT 300
241	TATTCATTCATTTGATTTGATTTGATTTGATTTGATTTGATTTGATTTGATTT 300
241	TATTCATTCATTTGATTTGATTTGATTTGATTTGATTTGATTTGATTTGATTT 300

Db 241 TAATGATTCATTTGGTTGGATTACACAGTTAAGTTTGTGTTCTTTGTTATATGT 300
QY 301 AATAGTAATCAAAACAAAGATATGATTAAGTGAACATATTTGTTATGACCCCA 360
Db 301 AATAGTAATCAAAACAAAGATATGATTAAGTGAACATATTTGTTATGACCCCA 360
QY 361 AAAAAAAAAAACAACAACAACACCCCGGATATATGTTTGTGTTGAT 420
Db 361 AAAAAAAAAAACAACAACAACACCCCGGATATATGTTTGTGTTGAT 420
QY 421 AGTTTATTTGATCATATTAACATGATTTCTTGTATTAATGATTAAGATTTCTTC 480
Db 421 AGTTTATTTGATCATATTAACATGATTTCTTGTATTAATGATTAAGATTTCTTC 480
QY 481 CAATTAATTTGATCATATTTCTGATTAATTAATTAATTAAGATGTAATATCC 540
Db 481 CAATTAATTTGATCATATTTCTGATTAATTAATTAATTAAGATGTAATATCC 540
QY 541 GTTTATGATCACTCCATCATATGATTAATGATTTCTGCTAATCCAGCAATTAATTA 600
Db 541 GTTTATGATCACTCCATCATATGATTAATGATTTCTGCTAATCCAGCAATTAATTA 600
QY 601 AGAGTTTGAAGAAAAACGAAAAATGAAGAAAGGAGTAGTACCATGAGATATG 660
Db 601 AGAGTTTGAAGAAAAACGAAAAATGAAGAAAGGAGTAGTACCATGAGATATG 660
QY 661 TGAATTAATTAATCAAGAGATTAAGATGACAAACAAAGTTGTAATATGTCCT 720
Db 661 TGAATTAATTAATCAAGAGATTAAGATGACAAACAAAGTTGTAATATGTCCT 720
QY 721 GCCAGTTTCTCTCAGCAATCATATGACCCATTTGATTTCTGATATTCGTTAAA 780
Db 721 GCCAGTTTCTCTCAGCAATCATATGACCCATTTGATTTCTGATATTCGTTAAA 780
QY 781 TTTGCAATACGATTTGAAAAATATTTATTTGATGATCTGATATTAATGTTCA 840
Db 781 TTTGCAATACGATTTGAAAAATATTTATTTGATGATCTGATATTAATGTTCA 840
QY 841 GGTATTTGATATATCTTCTGTTTAAAGCATATTTCTCTTTTGTGTTCTCTCT 900
Db 841 GGTATTTGATATATCTTCTGTTTAAAGCATATTTCTCTTTTGTGTTCTCTCT 900
QY 901 TAATTAATTAATTAATCGCGATATATGATTAACAATGATATTCACAAACATTTGTC 960
Db 901 TAATTAATTAATTAATCGCGATATATGATTAACAATGATATTCACAAACATTTGTC 960
QY 961 ACCATTTGAATTAATTTTCTCAACATTTACGGACACTGACCTGACCCCTTAATA 1020
Db 961 ACCATTTGAATTAATTTTCTCAACATTTACGGACACTGACCTGACCCCTTAATA 1020
QY 1021 CGATTTTACAGCGTCACTAGTTGAGATTAATGATTAAGCATTAAGCATTAAGCAT 1080
Db 1021 CGATTTTACAGCGTCACTAGTTGAGATTAATGATTAAGCATTAAGCATTAAGCAT 1080
QY 1081 TATTTATACAAAGTTTCAAACTGAAATATGCTTGAATTCCTTGAATAATTTGAA 1140
Db 1081 TATTTATACAAAGTTTCAAACTGAAATATGCTTGAATTCCTTGAATAATTTGAA 1140
QY 1141 CCGGTTGTTATGTAATTAATGATTTAGTGTAAACAATATGTTATCAATTAATG 1200
Db 1141 CCGGTTGTTATGTAATTAATGATTTAGTGTAAACAATATGTTATCAATTAATG 1200
QY 1201 ACATATACATTAATTTCTTACAGAAAAACAACTTAAGAGAGTTAAGCATATCCATAT 1260
Db 1201 ACATATACATTAATTTCTTACAGAAAAACAACTTAAGAGAGTTAAGCATATCCATAT 1260
QY 1261 GGGTATGCTATACCTTTCACTGATGCTATACAGACTTAAGAAATAGTATGATGTC 1320
Db 1261 GGGTATGCTATACCTTTCACTGATGCTATACAGACTTAAGAAATAGTATGATGTC 1320
QY 1321 GATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 1380
Db 1321 GATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 1380

QY 1381 AATCATTTCTTGGTGTCAACAATAAAAAACAAACAAAGAAAAACGATTTT 1440
Db 1381 AATCATTTCTTGGTGTCAACAATAAAAAACAAACAAAGAAAAACGATTTT 1440
QY 1441 CTGATGATTCATTCATGATGATTAATTAATTAATTAATTAATTAATTAATTA 1500
Db 1441 CTGATGATTCATTCATGATGATTAATTAATTAATTAATTAATTAATTAATTA 1500
QY 1501 TCTACAGCGTGAACCATATGCAATTAATTAATTAATTAATTAATTAATTAATTA 1560
Db 1501 TCTACAGCGTGAACCATATGCAATTAATTAATTAATTAATTAATTAATTAATTA 1560
QY 1561 ATTATTTGTTAGTTGAATTAATTAATTAATTAATTAATTAATTAATTAATTA 1620
Db 1561 ATTATTTGTTAGTTGAATTAATTAATTAATTAATTAATTAATTAATTAATTA 1620
QY 1621 ATACTAGTATGATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 1680
Db 1621 ATACTAGTATGATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 1680
QY 1681 TCTACAGCTTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 1740
Db 1681 TCTACAGCTTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 1740
QY 1741 ATGAAGATTAACCTTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 1800
Db 1741 ATGAAGATTAACCTTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 1800
QY 1801 AATTACCGTATGTTTGTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTT 1860
Db 1801 AATTACCGTATGTTTGTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTT 1860
QY 1861 TCTGCAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 1920
Db 1861 TCTGCAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 1920
QY 1921 ACATTTTAAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 1980
Db 1921 ACATTTTAAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 1980
QY 1981 AATGATGCTTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 2040
Db 1981 AATGATGCTTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 2040
QY 2041 TAAATACATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 2100
Db 2041 TAAATACATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 2100
QY 2101 AAATGACAAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 2160
Db 2101 AAATGACAAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 2160
QY 2161 TTGTTAAGCAAACTTCCAAATTTTATTTTATGGAACAAACAAATTAATTAATTA 2220
Db 2161 TTGTTAAGCAAACTTCCAAATTTTATTTTATGGAACAAACAAATTAATTAATTA 2220
QY 2221 CTATTTTGTGTTGATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 2280
Db 2221 CTATTTTGTGTTGATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 2280
QY 2281 AGCTATACGCGTCAAAAGTATTAATTAATTAATTAATTAATTAATTAATTAATTA 2340
Db 2281 AGCTATACGCGTCAAAAGTATTAATTAATTAATTAATTAATTAATTAATTAATTA 2340
QY 2341 TCAAGATTTGGGACAAACAAATTAATTAATTAATTAATTAATTAATTAATTAATTA 2400
Db 2341 TCAAGATTTGGGACAAACAAATTAATTAATTAATTAATTAATTAATTAATTAATTA 2400
QY 2401 TTGAGTAAATGTTTCTGATTAATTAATTAATTAATTAATTAATTAATTAATTA 2460
Db 2401 TTGAGTAAATGTTTCTGATTAATTAATTAATTAATTAATTAATTAATTAATTA 2460

Dh 961 ACCATTTTGATTAACCTTTTCTCMAACATTCGGGACACTGGACCTCGACCTTTAAATA 1020
Qy 1021 CGATTTTACAGCTCACTAGTGTAGATTAAGCATTAAGCAATTAAGAGCCGCTCAAGC 1080
Dh 1021 CGATTTTACAGCTCACTAGTGTAGATTAAGCATTAAGCAATTAAGAGCCGCTCAAGC 1080
Qy 1081 TATTTATACAAAGTACAAACGTGAATATACCTTGAAATCTTTAGAAAATTTTGAATTA 1140
Dh 1081 TATTTATACAAAGTACAAACGTGAATATACCTTGAAATCTTTAGAAAATTTTGAATTA 1140
Qy 1141 CGGTTGTTTATTAATATATAGATTTAGTGTAAACAAATATGTTAAATCAATTAAGTGTCA 1200
Dh 1141 CGGTTGTTTATTAATATATAGATTTAGTGTAAACAAATATGTTAAATCAATTAAGTGTCA 1200
Qy 1201 ACATATACATTAATTCCTTACAGAAAAACAACCTTAAGAGAGTAAATATCATATATAT 1260
Dh 1201 ACATATACATTAATTCCTTACAGAAAAACAACCTTAAGAGAGTAAATATCATATATAT 1260
Qy 1261 GGGTATGCTATACCTTTACAGTATGCTATATAGAGACTTAAGAAATATGTTATGATGTC 1320
Dh 1261 GGGTATGCTATACCTTTACAGTATGCTATATAGAGACTTAAGAAATATGTTATGATGTC 1320
Qy 1321 GATTAATGAAATTCACAGCGGTGTAATATATGAGACCGTATGTTAGCATCTGCA 1380
Dh 1321 GATTAATGAAATTCACAGCGGTGTAATATATGAGACCGTATGTTAGCATCTGCA 1380
Qy 1381 ATATCATTTCTGGTGGTCAACATTAACAAACAAACAAACAAACAAACAAACAAACAAAC 1440
Dh 1381 ATATCATTTCTGGTGGTCAACATTAACAAACAAACAAACAAACAAACAAACAAACAAAC 1440
Qy 1441 CTGGAATTCATTCATGATGATTAATGATGATCTTTGGGTACAGTTTGGAAGTCC 1500
Dh 1441 CTGGAATTCATTCATGATGATTAATGATGATCTTTGGGTACAGTTTGGAAGTCC 1500
Qy 1501 TCTACAGCGGTGTAACCATTCGACATTAATTAATGCTTTCTTAATGATCTTTAAAT 1560
Dh 1501 TCTACAGCGGTGTAACCATTCGACATTAATTAATGCTTTCTTAATGATCTTTAAAT 1560
Qy 1561 ATTTATTTGTTAGTTGAATTAATTAAGAGCAATGTAACATTAATTAATTAATTAAT 1620
Dh 1561 ATTTATTTGTTAGTTGAATTAATTAAGAGCAATGTAACATTAATTAATTAATTAAT 1620
Qy 1621 ATATCATTTGATGATTAATTCACAAATATCTTTGATGTTTAACTTAATCTTTGTTCT 1680
Dh 1621 ATATCATTTGATGATTAATTCACAAATATCTTTGATGTTTAACTTAATCTTTGTTCT 1680
Qy 1681 TCTACAGGTATTAATATTAATCATGAGGTAAAGAAAGTTTGTCTTAATTTGCGCATGC 1740
Dh 1681 TCTACAGGTATTAATATTAATCATGAGGTAAAGAAAGTTTGTCTTAATTTGCGCATGC 1740
Qy 1741 ATGAAGGATTAACCTTAATGATCTTTTGAATGTAATGTAATGTAATGTAATGTAAT 1800
Dh 1741 ATGAAGGATTAACCTTAATGATCTTTTGAATGTAATGTAATGTAATGTAATGTAAT 1800
Qy 1801 AATTAACCGTATGTTTGTGTCATATGACAGCCCTTACACTGATGATGTAATTTT 1860
Dh 1801 AATTAACCGTATGTTTGTGTCATATGACAGCCCTTACACTGATGATGTAATTTT 1860
Qy 1861 TCTGCAAAATTAATTAAGAAATTCATGCTATCAATAGAAAGAAAGAGTAT 1920
Dh 1861 TCTGCAAAATTAATTAAGAAATTCATGCTATCAATAGAAAGAAAGAGTAT 1920
Qy 1921 ACATTTTAAATTAAGACAAATTTTGAAGAAATGTAATTAATTTTACAAATTTTAA 1980
Dh 1921 ACATTTTAAATTAAGACAAATTTTGAAGAAATGTAATTAATTTTACAAATTTTAA 1980
Qy 1981 AATATATGCTTATATATATTTCTTAAATATTTTATATATTTATATTTT 2040
Dh 1981 AATATATGCTTATATATATTTCTTAAATATTTTATATATTTATATTTT 2040
Qy 2041 TAAATCATTAATGAAACCAATTAATAGTTGGTGAATTCAAATATCTCCATTAATTTT 2100
Dh 2041 TAAATCATTAATGAAACCAATTAATAGTTGGTGAATTCAAATATCTCCATTAATTTT 2100

Qy 2101 AAATCTCAAAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 2160
Dh 2101 AAATCTCAAAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 2160
Qy 2161 TTGTTAAACAGAACTTCCAAATTTTATTTTATGACAGCAAGAAATTAACAGATAGAAA 2220
Dh 2161 TTGTTAAACAGAACTTCCAAATTTTATTTTATGACAGCAAGAAATTAACAGATAGAAA 2220
Qy 2221 CTATTTGTTGTTGGAAGTAAGTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 2280
Dh 2221 CTATTTGTTGTTGGAAGTAAGTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 2280
Qy 2281 AGCCTATACCGCTCAAGATATGTTATCTAGTATGTTATTAATTAATTAATTAATTAAT 2340
Dh 2281 AGCCTATACCGCTCAAGATATGTTATCTAGTATGTTATTAATTAATTAATTAATTAAT 2340
Qy 2341 TCAAGATTTGGACAAACATTAAGAAACGGAATTTAAATTAATTAATTAATTAATTAAT 2400
Dh 2341 TCAAGATTTGGACAAACATTAAGAAACGGAATTTAAATTAATTAATTAATTAATTAAT 2400
Qy 2401 TTGAGTAAATGTTGTTTCTGATATGAGGGGCAAAAAAGCAATGCCAAAGCTAC 2460
Dh 2401 TTGAGTAAATGTTGTTTCTGATATGAGGGGCAAAAAAGCAATGCCAAAGCTAC 2460
Qy 2461 GGGTGTGACCTGCTCAGTTCCGTATTAATTAATTAATTAATTAATTAATTAATTAATTA 2520
Dh 2461 GGGTGTGACCTGCTCAGTTCCGTATTAATTAATTAATTAATTAATTAATTAATTAATTA 2520
Qy 2521 TAGGGTCTCTTGACATTTTCTACGTTTCTACCCCTACTGTGAGCCGACCTTTTCCCA 2580
Dh 2521 TAGGGTCTCTTGACATTTTCTACGTTTCTACCCCTACTGTGAGCCGACCTTTTCCCA 2580
Qy 2581 TATCTTAAGGGTATTTTGGAAATCCCAATTTAAACGATTTGAGACCGTACCGGACTTC 2640
Dh 2581 TATCTTAAGGGTATTTTGGAAATCCCAATTTAAACGATTTGAGACCGTACCGGACTTC 2640
Qy 2641 TGGGATTCCTGAGACATTTTCAAAAAATTAATTAAGCAGAAATGGTTTATTAATTTAA 2700
Dh 2641 TGGGATTCCTGAGACATTTTCAAAAAATTAATTAAGCAGAAATGGTTTATTAATTTAA 2700
Qy 2701 AACTCAACATTTGATGATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 2760
Dh 2701 AACTCAACATTTGATGATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 2760
Qy 2761 ATCTAATGATTTTCTTCTACAGGGGATGAAGTATTAATTAATTAATTAATTAATTAAT 2820
Dh 2761 ATCTAATGATTTTCTTCTACAGGGGATGAAGTATTAATTAATTAATTAATTAATTAAT 2820
Qy 2821 ACAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 2880
Dh 2821 ACAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 2880
Qy 2881 CTATTTGATGATGATGATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 2940
Dh 2881 CTATTTGATGATGATGATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 2940
Qy 2941 AAGTATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 3000
Dh 2941 AAGTATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 3000
Qy 3001 TGAAGCTCATTTGTTAGTTTAACTTAATTAATTAATTAATTAATTAATTAATTAATTA 3060
Dh 3001 TGAAGCTCATTTGTTAGTTTAACTTAATTAATTAATTAATTAATTAATTAATTAATTAAT 3060
Qy 3061 ATTAACATTAATTAATTTCTGATCTTTTAAACCCCTTACAAACGAAAGCTCCCTTT 3120
Dh 3061 ATTAACATTAATTAATTTCTGATCTTTTAAACCCCTTACAAACGAAAGCTCCCTTT 3120
Qy 3121 TCAGTAAAGTCCGATTTCCCAATCTTAAGACAAAGCCATTTAGAAAGAGAAAGTGA 3180
Dh 3121 TCAGTAAAGTCCGATTTCCCAATCTTAAGACAAAGCCATTTAGAAAGAGAAAGTGA 3180

RESULT 4
US-10-312-841-1
: Sequence 1, Application US/10312841
: GENERAL INFORMATION:
: APPLICANT: Epigenomics AG
: TITLE OF INVENTION: Diagnose von bedeutenden genetischen Parametern innerhalb des MHC
: FILE REFERENCE: E01/1208/WO
: CURRENT APPLICATION NUMBER: US/10/312, 841
: NUMBER OF SEQ ID NOS: 2
: SEQ ID NO 1
: LENGTH: 3673778
: TYPE: DNA
: ORGANISM: Artificial Sequence
: FEATURE:
: OTHER INFORMATION: chemically treated genomic DNA (Homo sapiens)
: NAME/KEY: unsure
: LOCATION: (3294164)
: US-10-312-841-1

3.2%: Score 103; DB 9; Length 3673778;

Best Local Similarity 42.4%; Pred. No. 6.1e-07;
Matches 989; Conservative 0; Mismatches 1320; Indels 24; Gaps 7;

QY 1 ATGTGGTATTAATTTGGTGGTTCGCTTACCAATTAATTAATTCGTTCTGCT 60
DB 1713239 ATTTATTTTATATATATATATATATATATATATATATATATATATG 1713298
QY 61 TATTCGTACATGATTTAGTTGGTCTCATTTGGATTCAGATATTAATTAAT 120
DB 1713399 TTTTAT 1713358
QY 121 AAAATCATTTAAATATTTTCAAGATTAATTAATTAATTCGTTATTAACAA 180
DB 1713359 TATATATATTTTATATATATATATATATATATATATATATATATAT 1713415
QY 181 AATATCTATCTTGGTATATGAGAAATATGAGTTGGAATTAATTAATTAAG 240
DB 1713416 TAT 1713475
QY 241 TAATCGATTCATTTGGTGGATTAACAGTTAAGTTTGTGTTCTGTTATATGT 300
DB 1713476 TTTATATATATTTTATATATATTTTATATATATTTTATATATATAT 1713535
QY 301 ATATGATAATCAAAAAGATATGATGAGTGAACATTAATTCGTTATGACCCA 360
DB 1713536 ATTTTATATATATTTTATATATATATATATATATATATATATATAT 1713595
QY 361 AAAAAAAGAAAAAACAACAAACCCCCCGGATATGTTTGGTCTGCGATT 420
DB 1713596 ATATATTTTATATATATTTTATATATTTTATATATTTTATATATAT 1713655
QY 421 AGGTTATTTGATCATTAATACATGATCATTTCTTTGATTACATGAATTTCTTAC 480
DB 1713656 TTTATATATATTTTATATATTTTATATATATTTTATATATATATATA 1713715
QY 481 CAAATTAATTTGCAATATCTCTGATATTAATTAATTAATACAGTGTGAATTC 540
DB 1713716 TAAATA 1713775
QY 541 GTTATTCGATCCTCCATCATGATATGATCTTGGCTAATCCGCAAAATTAACA 600
DB 1713776 TATTTATATATATTTTATATATTTTATATATTTTATATATATTTAT 1713835
QY 601 AGAGTATGAGAAAAACGAAATTAAGAAAGGAAGAGTGTGAGCCATGGAGATG 660
DB 1713836 TATATATTTTATATATATATATATTTTATATATATTTTATATATATAT 1713895
QY 661 TGAATATATTAACAAGAGATA-AGAGATGACAACAAAGTTGGAATATGTCGC 719
DB 1713896 TTTATATATATTTTATATATATTTTATATATATTTTATATATATTTTAT 1713955

QY 720 TGCCAGCTTCTCTACAAATCAATATGACCCATTTGGATTTTCTGGATATTCGTTAAA 779
DB 1713956 ATATATATTTTATATATATTTTATATATATTTTATATATATTTTATATATATTTTAT 1714015
QY 780 ATTTGGCAATACGATTTGCAAAAAATTTTATTTTGTGATGATCTCAATATTTGTTCC 839
DB 1714016 TATATATTTTATATATATTTTATATATATTTTATATATATTTTATATATATTTTAT 1714075
QY 840 AGGATTTTGCATATCTCTGTTTAAAGCATATTTGCTTTTGTGTTGCTCTC 899
DB 1714076 ATATATTTTATATATATTTTATATATTTTATATATATTTTATATATATTTTATATA 1714135
QY 900 TTAACATATATATATCGCGATATATGATTAACAATGATATATCAAAACAAATGCTCG 959
DB 1714136 TATATTTTATATA-----TATATTTTATATATATTTTATATATATTTTATATATTTTAT 1714189
QY 960 GACATTTTGAATTAACCTTTTTCGCAACATTTACGGACACTGACCTGACCTTAAT 1019
DB 1714190 ATATATATTTTATATATTTTATATATATTTTATATATATTTTATATATTTTATATA 1714249
QY 1020 ACGATTTTACAGCTCATGATGAGATTAATCTACATATA---AAGCATTAAGACCGCTTC 1076
DB 1714250 TTTTATATATATTTTATATATTTTATATATATTTTATATATATTTTATATATTTTAT 1714309
QY 1077 AAGCTATTTATACAACTTACAAACGTAATATAGCTTGAATTCCTTAGAAATTTTGA 1136
DB 1714310 TATATATTTTATATATATTTTATATATTTTATATATATTTTATATATATTTTATAT 1714369
QY 1137 ATTAACGGTGTATATGTAATATATATATTTAGTGTAAACAATATGTAATCAATATAG 1196
DB 1714370 ATATATATTTTATATATATTTTATATATATTTTATATATATTTTATATATATTTTAT 1714429
QY 1197 GTCACATATATACATATATCTTACAGAAAAACAACCTTAAAGAGATTAACATATCCAT 1256
DB 1714430 TATATATTTTATATATATTTTATATATTTTATATATATTTTATATATATTTTATAT 1714489
QY 1257 ATATGGTATGCTATACCTTTCACGATGCTATATCTAGAGCTTAAGAAATGTTATGCTGA 1316
DB 1714490 ATATATTTTATATATATTTTATATATTTTATATATATTTTATATATATTTTATAT 1714549
QY 1317 TGTGCAATTAATGAATTCACACGCGGTGATATATTTGAGACCGCTGATGATCAGT 1376
DB 1714550 TATATATTTTATGATATATATTTTATATATATTTTATATATATTTTATATATTTTAT 1714602
QY 1377 GCAATATATCATTTCTTGGTGTGCAACATTAATAAACAACAAAAAGAAAGACAT 1436
DB 1714603 ATATATATTTTATATATATTTTATATATATTTTATATATATTTTATATATATTTTAT 1714662
QY 1437 TTTTCTGGATTCATTCATATGATCTTAATATGATATGATCTTTGGGTACAGTTGCA 1496
DB 1714663 ATATTTTATATATATTTTATATATTTTATATATTTTATATATTTTATATATTTTAT 1714722
QY 1497 GTCCTACAAAGGTGTAACATCTGCACATTAATTAATTTCTTTAATGATCTTTA 1556
DB 1714723 TGTATATTTTATATATATTTTATATATTTTATATATTTTATATATTTTATATATAT 1714782
QY 1557 ACAATATTTATGTTAGTTGGAATTTAATAAGACGAACCTGTACATTAACATATTTATA 1616
DB 1714783 TTAATATATTTTATATATATTTTATATATTTTATATATATTTTATATATATTTTATATA 1714842
QY 1617 TTGATATCTAGTATGATATTTTCCAAATACATATCTTTGGATGTTTAACTTAATCTTGT 1676
DB 1714843 TATATATTTTATATATATTTTATATATTTTATATATTTTATATATATTTTATATATATA 1714902
QY 1677 TTTCTCTACGATATTAATTTTATATCTTCAGGTAAGAAAAAGTTTCTTATTTTGGCG 1736
DB 1714903 TTTTATATATATTTTATATATATTTTATATATATTTTATATATATTTTATATATTTTAT 1714962
QY 1737 ATGCATGAGGATTAACCTTAATGATCTTAATTTTGTGAAAA---TGTAACTTTTATCTC 1793
DB 1714963 ATATATATTTTATATATATTTTATATATTTTATATATATTTTATATATATTTTATAT 1715022


```

QY 1514 AACCATCTGCACCTTTAAATGCTTCTTTAATGCACTTTAAACATTTATTTAGTACT 1573
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 134169 TAAATATA--GAATATTAATATGATATTAATCAATATTAATTAATTAATTA 134113
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 1574 TGGATTTAATAGGCGCACTGTACATTTACATTTATTTATTTAGTACTAGATG 1633
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 134112 TTTAATATATGATATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 134053
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 1634 ATTTATCCAAATACATCTTGTGATGTTTAACTTAACCTTCTTCTTCCGCGGATTA- 1692
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 134052 TTATTTTATATACCTATTATTAACATTAATTAATTAATTAATTTATTTATTTATTC 133993
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 1693 -AAATTAATCAATGAGGTAATAAAAGTTTGTCTTAATTTGCGCATGATGAGAGATA 1751
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 133992 TAAATTTAATATTAATTAATAAATAATTAATTAATTAATTAATTAATTAATTAATTA 133933
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 1752 AACTTAATGACTTAATTTTGTGAATAATGAACCTTTTACATGATTAATTTACCGTAT 1811
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 133932 AATTTATATTAATTAATAAAATAATTAATTAATTAATTAATTAATAAATAATTAATTAAT 133873
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 1812 GTTTTGTGCGCATTAATGACAGCTCTACCACTGTGATGATGCAATTTTCTGCAATAT 1871
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 133872 AAATTAATATGATTTATTAATTAATTAATTAATTAATTAATTAATTTATTTATTTATTC 133813
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 1872 ----TAAATTAAGATTCATGCTACTATCAATAGAGAAACAGCTGATTTTACATTT 1926
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 133812 ATTTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 133753
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 1927 TAAATTAAGCAAAATTTTGAATAATGTTATTTCTTACATTTATTTAAATATG 1986
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 133752 ATATTTAATTAATTAATTTTATTAATAATTAATTTATTAATTAATTAATTAATTAAT 133693
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 1987 ATGCTTAATTAATTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCT 2046
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 133692 TATATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 133633
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 2047 CATTTGAACCAATTAATGTTGGTGAATTCAAATATCCCATTAATTTTGAATCT 2106
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 133632 TA-AATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 133574
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 2107 ACAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 2166
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 133573 ATTTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 133514
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 2167 ACAGAACTTCCAAATTTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCT 2226
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 133513 ATATTTACTTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 133454
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 2227 TGTGTGCAATGGAAGTATGATTAATTAATTAAGCAAAATTTTAAAAATTAATTAAGCCTA 2286
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 133453 TTTCTTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 133394
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 2287 TACGCGCTCAAGATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 2313
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 133393 TATATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 133367
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

```

RESULT 7
US-09-947-911-2/c

```

; Sequence 2, Application US/09947911
; GENERAL INFORMATION:
; APPLICANT: VENTER, J. Craig
; TITLE OF INVENTION: COLLECTION OF SINGLE NUCLEOTIDE POLYMORPHISM (SNPs) LOCATED ON CH
; TITLE OF INVENTION: X, METHODS OF DETECTION, AND USES THEREOF
; FILE REFERENCE: C1001300
; CURRENT FILING DATE: 2001-09-07
; NUMBER OF SEQ ID NOS: 368
; SEQ ID NO 2
; LENGTH: 600322
; TYPE: DNA
; ORGANISM: HUMAN
; FEATURE:

```

```

; NAME/KEY: misc_feature
; LOCATION: (1)...(600322)
; OTHER INFORMATION: n = A,T,C or G
US-09-947-911-2

Query Match      3.1%; Score 99.4; DB 7; Length 600322;
Best Local Similarity 44.6%; Pred. No. 2.8e-06;
Matches 735; Conservative 0; Mismatches 891; Indels 21; Gaps 8;

QY 676 GAGAAATAGAGATGACACCAAAAGGTTGGAATTAATGTCCTGCGCATTTCTGCA 735
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 501609 GACAAACCAATAGGAATTTATGCGGTTGGAGCTGCGCTCTGCTCCAGCCTGCA 501550
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 736 CAATCAATGACACCTATTTGATTTTCTGATATTTCTTAATTTTCCGATTAAGAT 795
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 501549 GAATATGATTTCTAAGTTCCACAGAGAGGATTCATGCTCAACAGACATTAATTAAT 501490
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 796 GTGAATAATATTTATTTTGTAGCTGATCTCAATATTAATGTCAGATTTTGCATATTC 855
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 501489 TATATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 501430
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 856 TTCTGTTAA-AGCATATTTTGTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCT 914
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 501429 AAATATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 501370
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 915 CGCGGATATGATTAACATGATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 974
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 501369 ATATGATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 501310
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 975 ACTTTTCTCAAACTTAACGAGCACTGACCTGACCTTAATTAATTAATTAATTAATTAATTAATTAAT 1034
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 501309 ACAATATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 501250
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 1035 CACTAGTTGAGAT-TACTAGCATTAAGCATTAAGCAACCGTTCAAGCTATTTATCAAG 1093
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 501249 TAAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 501190
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 1094 TTCAACAGCATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 1153
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 501189 TATTAACATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 501130
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 1154 AAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 1213
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 501129 AATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 501071
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 1214 TCCTTAAGCAAAACCAACTTAAGAGAGTTTACATATCCATTAATGAGTATGCTATTC 1273
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 501070 TCAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 501011
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 1274 CTTCAGCATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 1333
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 501010 TTATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 500951
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 1334 CACACGCGTGAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 1393
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 500950 AATA-----TTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 500898
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 1394 TTGCTCAACATTAATAAACAAGAAAGAAAGAAAGCAATTTTCTTGATTCATTT 1453
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 500897 ATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 500838
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 1454 CAATGATTAATAATGATAGATCTTTGGGTAGAGTTTTCGAAAGCTCTCAACAGCGT 1513
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 500837 ACATGATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 500778
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 1514 AACCATCTGCACATTAATTAATTTCTTTAATGCACTTTAACAATTTATTTAGTATG 1573
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 500777 TAAATATA--GAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 500721
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 1574 TGGATTTAATTAAGGCGCACTGTGAATTAATTAATTTATTTAGTACTAGATG 1633
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 500720 TTTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 500661
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

```

QY 1634 ATTATTCACATACACTTGGATGTTAACTTAATCTGTTCTTCTACGGATA- 1692
 Db 500660 TTATTTTATATCTTATTTTATACATTAATATAATTTATTTTATATATAC 500601
 QY 1693 -AATATATATCATGAGGTAAAAAGTTTGTCTTATTTTCGGATGATGAGATTA 1751
 Db 500600 TAAATTTTATATATTTTATATATATTTATATATATATATATATATATATAT 500541
 QY 1752 ACCATAGACTTAAATTTTGAATAATGTAACCTTTTACTCATATATATACCGTA 1811
 Db 500540 AATTTAT 500481
 QY 1812 GTTTTGTTCCTAATAGACGCTACACAGCTGTGATCTCATTTTCTCGCAATAT 1871
 Db 500480 AAT 500421
 QY 1872 -TAAATTAGAATTCATCTACTCATATAGAGAAACGCTGATATATACATTT 1926
 Db 500420 ATTTAT 500361
 QY 1927 TAAATTAAGACAAAATTTTGAATAATGTTATATTTCTACAAATATATATATAT 1986
 Db 500360 AT 500301
 QY 1987 ATGCTATATATGATTTCCATGCTCTTAAATATTTTATATATATATATATAT 2046
 Db 500300 TAT 500241
 QY 2047 CATATAGAACCAAT 2106
 Db 500240 TAT 500182
 QY 2107 ACAAT 2166
 Db 500181 ATTTAT 500122
 QY 2167 ACAGAACTCCAAATTTTCTTATATGACAAAGAAATATACAGATAGAAACATTT 2226
 Db 500121 AT 500062
 QY 2227 TGTGTGATGAGATGATATATATATATATATATATATATATATATATATAT 2286
 Db 500061 TTTTAT 500002
 QY 2287 TACGCGCTCAAGTATGATCTAGTA 2313
 Db 500001 TAT 499975

 RESULT 8
 US-09-947-911-331
 : Sequence 331, Application US/09947911
 : GENERAL INFORMATION:
 : APPLICANT: VENTER, J. Craig
 : TITLE OF INVENTION: COLLECTION OF SINGLE NUCLEOTIDE POLYMORPHISM (SNPs) LOCATED ON CH
 : FILE REFERENCE: CL001300
 : CURRENT APPLICATION NUMBER: US/09/947, 911
 : CURRENT FILING DATE: 2001-09-07
 : NUMBER OF SEQ ID NOS: 368
 : SEQ ID NO 331
 : LENGTH: 2453797
 : TYPE: DNA
 : ORGANISM: HUMAN
 : FEATURE:
 : NAME/KEY: misc_feature
 : LOCATION: (1)-(2453797)
 : OTHER INFORMATION: n = A,T,C or G
 US-09-947-911-331

 Query Match 3.0%; Score 97.4; DB 7; Length 2453797;
 Best Local Similarity 44.9%; Pred. No. 4.3e-06;
 Matches 469; Conservative 0; Mismatches 561; Indels 15; Gaps 2;

QY 1407 AAAACAAAACAGAAAAAGAAAGATTTTCTTGATTCATTCATGATCTAAA 1466
 Db 186415 ATAAATATAGAT 186474
 QY 1467 TGCATATCTTTTGGGTACGTTTGGAAAGTCCCTACACAGCTGTACCATCTGCAC 1526
 Db 186475 TGTACAT 186534
 QY 1527 TATTAATTCCTTCTTAAATGATCTTAAACAT----ATTATGTTAGTGGAAATTA 1582
 Db 186535 GTATTAAT 186594
 QY 1583 ATAGAGCGAATCTTACATTTACATATATATATATATATATATATATATATAT 1642
 Db 186595 TATTAAT 186654
 QY 1643 AATACATCTTGGATCTTTAACTTAACTTATCTGTTCTTCTCCAGGATATATATAT 1702
 Db 186655 GAT 186714
 QY 1703 ATGAGGTAAAGAAAGTTTGTCTTATTTTCGGATGATGAGATTAACCTATGACT 1762
 Db 186715 AT 186774
 QY 1763 TTAATTTTGAAGATGTAACCTTTTACTCATATATATATATATATATATATATAT 1822
 Db 186775 AAT 186834
 QY 1823 CATATAGACAGCTCTACAACTGTGATGATGATATTTTCTGCAATATATATATAT 1882
 Db 186835 AT 186894
 QY 1883 TTCAATGCTCTATCATATAGAGAAACAGCTGATATATATATATATATATATAT 1942
 Db 186895 TAAAT 186954
 QY 1943 TTTTGAAGAAAGTTATATATTTCTACAAATATATATATATATATATATATAT 2002
 Db 186955 TAT 187003
 QY 2003 TCCATGCTCTTAAATATTTTATATATATATATATATATATATATATATATAT 2062
 Db 187004 AT 187063
 QY 2063 TACTGTGATTCATCAATATCTCATATATATATATATATATATATATATATATAT 2122
 Db 187064 TCAAT 187123
 QY 2123 TACTCATTAACAATGATAGAGAAAGTCCAAAAAATTTGTTAACGAAACTTCAAT 2182
 Db 187124 ATAAAT 187183
 QY 2183 TTTTATTTTATATAGAACAGAAATACAGATAGAAACATATTTTGTGGAATGGA 2242
 Db 187184 ATAAAT 187243
 QY 2243 TACTAAT 2302
 Db 187244 TTAAT 187303
 QY 2303 GTTATCTAGTAGTGAATTAATATATATATATATATATATATATATATATATAT 2362
 Db 187304 ATAAAT 187363
 QY 2363 AACGGAATTAAT 2422
 Db 187364 TAT 187423
 QY 2423 TATTGAGGGCAAAAAAGACAAAT 2447
 Db 187424 ACAT 187448

Query Match	2.98;	Score 94.2;	DB 7;	Length 597573;
Best Local Similarity	45.28;	Pred. No. 1.6e-05;		
Matches 519; Conservative	0;	Mismatches 608;	Indels 20;	Gaps 4;

[illegible]

Wed Jun 11 10:32:17 2003

us-09-502-426b-1_copy_1_3202.rnpn

Page 16

Dd 1092021 ATAAATATTATTAATTTAATAATTAATATTTAATATAATTAATATAATTTATATAT 1091962

Gy 2275 TATATAAGCCATACGGCGCTCAAGTATGTTATCTAGTAGTGTAATTAATGCAT 2332
||||| - |||| - |||| - |||| - ||||
Db 1091961 TATATATGTATATTTATATATATATATAAAAAGAAATATATATTTATATATCAT 1091904

Search completed: June 11, 2003, 07:57:02
Job time : 2411.47 secs

•

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: June 10, 2003, 22:04:49 ; Search time 4854.18 Seconds
(without alignments)
10683.154 Million cell updates/sec

Title: US-09-502-426b-1_COPY_1_3202

Perfect score: 3202

Sequence: 1 atgtgggtattatattgtt.....gagagagagaactagctcc 3202

Scoring table: IDENTITY_NUC

Gapop 10.0 , Gapept 1.0

Searched: 16154066 seqs, 8097743376 residues

Total number of hits satisfying chosen parameters: 32308132

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

EST:*
1: em_estba:*
2: em_esthum:*
3: em_estin:*
4: em_estmu:*
5: em_estrov:*
6: em_estrpl:*
7: em_estro:*
8: em_hic:*
9: gb_est1:*
10: gb_est2:*
11: gb_hic:*
12: gb_est3:*
13: gb_est4:*
14: gb_est5:*
15: em_estfun:*
16: em_estom:*
17: gb_gss:*
18: em_gss_hum:*
19: em_gss_inv:*
20: em_gss_pln:*
21: em_gss_vrt:*
22: em_gss_fun:*
23: em_gss_mam:*
24: em_gss_mus:*
25: em_gss_other:*
26: em_gss_pro:*
27: em_gss_rtd:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	195	6.1	473	AL768954	Arabidops
2	148	4.6	148	BH811044	BH811044 Arabidops
3	146.4	4.6	473	AL768954	SAUK_0571
4	113.2	3.5	1101	CNS00EVL	AL069706 Drosophila
5	110.8	3.5	1101	CNS00EVL	AL069706 Drosophila
6	109.6	3.4	1101	CNS0021J	AL061936 Drosophila

C	7	98.4	3.1	836	17	CNS01100	AL099642 Drosophila
C	8	98.2	3.1	1101	17	CNS000807	AL069440 Drosophila
C	9	98	3.1	807	10	AV717372	AV717372 AV717372
C	10	98	3.1	1092	17	CNS020K7	AL175696 Tetradon
C	11	94	2.9	1101	17	CNS003BD	AL064091 Drosophila
C	12	94	2.9	1201	17	CNS0167M	AL06396 Drosophila
C	13	93.6	2.9	1101	17	CNS003BD	AL064091 Drosophila
C	14	93.6	2.9	1101	17	CNS000807	AL069440 Drosophila
C	15	93	2.9	1101	17	CNS00801	AL057419 Drosophila
C	16	92	2.9	661	17	CNS020VJ	AL209800 Tetradon
C	17	91.8	2.9	886	17	BH177277	BH177277 008_L_22-
C	18	91.8	2.9	886	17	CNS07JUX	AL161425 T3 end of
C	19	91.8	2.9	1101	17	CNS0039G	AL063921 Drosophila
C	20	90.6	2.8	945	17	CNS04D0K	AL285149 Tetradon
C	21	90.6	2.8	1187	17	BH1102	BH1102 F19C22-T7 I
C	22	90.4	2.8	836	17	CNS01100	AL099642 Drosophila
C	23	89.6	2.8	1029	17	CNS012GM	AL174271 Tetradon
C	24	89	2.8	987	17	CNS014PQ	AL104456 Drosophila
C	25	88.6	2.8	1092	17	CNS020K7	AL175696 Tetradon
C	26	88.6	2.8	1190	17	CNS020N7	AL206908 Tetradon
C	27	88.4	2.8	1225	17	CNS0161D	AL06171 Drosophila
C	28	88.2	2.8	1101	17	CNS0039G	AL063921 Drosophila
C	29	87.4	2.7	1169	17	CNS006KHQ	AL042900 T3 end of
C	30	87	2.7	1225	17	CNS0161D	AL106171 Drosophila
C	31	86.8	2.7	1101	17	CNS0021J	AL061936 Drosophila
C	32	86.6	2.7	866	17	AQ258984	AQ258984 nbx0021P
C	33	86	2.7	1013	17	CNS006RPQ	AL12260 T7 end of
C	34	85.8	2.7	1187	17	BH1102	BH1102 F19C22-T7 I
C	35	85.6	2.7	961	17	CNS008KH	AL051882 Drosophila
C	36	85.2	2.7	928	17	CNS0008X	AL071865 Drosophila
C	37	85	2.7	1101	17	CNS05AB2	AL125829 Tetradon
C	38	84.8	2.6	1084	17	CNS071NH	AL425139 clone BA0
C	39	84.8	2.6	1101	17	CNS003BB	AL064089 Drosophila
C	40	84.8	2.6	1190	17	CNS020N7	AL206908 Tetradon
C	41	84.6	2.6	804	17	B12681	B12681 F27D1-Sp6.1
C	42	84.4	2.6	1013	17	CNS006RPQ	AL142260 T7 end of
C	43	84.4	2.6	1101	17	CNS00801	AL057419 Drosophila
C	44	83.8	2.6	609	17	CNS025K2	AL182171 Tetradon
C	45	83.8	2.6	886	17	BH177277	BH177277 008_L_22-

ALIGNMENTS

RESULT 1	AL768954	LOCUS	Arabidopsis thaliana T-DNA flanking sequence GK-082A08-011867,
DEFINITION	AL768954	DEFINITION	Arabidopsis thaliana T-DNA flanking sequence GK-082A08-011867,
ACCESSION	AL768954	ACCESSION	AL768954
VERSION	AL768954.1	VERSION	AL768954.1
KEYWORDS	GSS.	KEYWORDS	GSS.
SOURCE	thale cress.	SOURCE	thale cress.
ORGANISM	Arabidopsis thaliana	ORGANISM	Arabidopsis thaliana
REFERENCE	1	REFERENCE	1
AUTHORS	Strizhov, N., Li, Y., Rosso, M., Viehoever, P., Dekker, K., Saedler, H. and Weishaar, B.	AUTHORS	Strizhov, N., Li, Y., Rosso, M., Viehoever, P., Dekker, K., Saedler, H. and Weishaar, B.
TITLE	A pipeline for automated high-throughput generation of FSTs (flanking sequence tags) from Arabidopsis thaliana T-DNA transformed lines	TITLE	A pipeline for automated high-throughput generation of FSTs (flanking sequence tags) from Arabidopsis thaliana T-DNA transformed lines
JOURNAL	Unpublished	JOURNAL	Unpublished
REFERENCE	2	REFERENCE	2
AUTHORS	Rosso, M., Strizhov, N., Li, Y., Reiss, B., Dekker, K. and Weishaar, B.	AUTHORS	Rosso, M., Strizhov, N., Li, Y., Reiss, B., Dekker, K. and Weishaar, B.
TITLE	A new Arabidopsis thaliana T-DNA mutagenised population (GABI-Kat) for flanking sequence tag based reverse genetics	TITLE	A new Arabidopsis thaliana T-DNA mutagenised population (GABI-Kat) for flanking sequence tag based reverse genetics
JOURNAL	Unpublished	JOURNAL	Unpublished
AUTHORS	3 (bases 1 to 473)	AUTHORS	3 (bases 1 to 473)
TITLE	Li, Y., Rosso, M., Strizhov, N. and Weishaar, B.	TITLE	Li, Y., Rosso, M., Strizhov, N. and Weishaar, B.
JOURNAL	Submitted (17-JUN-2002) Weishaar B., Max-Planck-Institut fuer	JOURNAL	Submitted (17-JUN-2002) Weishaar B., Max-Planck-Institut fuer

TITLE Direct Submission
JOURNAL Submitted (17-JUN-2002) Weissnaar B., Max-Planck-Institut fuer Zuechtungsforchung, Carl-von-Linne-Weg 10, Koeln, 50829, Germany
COMMENT This sequence is recovered from the left border of the T-DNA. It indicates an insertion within the locus defined by clone T3A5. The sequences are generated at the MPI for Plant Breeding Research in the context of the GABI-Kat project. GABI-Kat is part of the German Plant Genomics program designated 'GABI'. Information on line availability can be found at: <http://www.mpiz-koeln.mpg.de/GABI-Kat/>.

FEATURES
source
1. 473
location/Qualifiers
/organism="Arabidopsis thaliana"
/strain="Columbia 0"
/db_xref="taxon:3702"
/clone="GK-082A08-011867"
/clone_lib="Arabidopsis thaliana T-DNA insertion lines"
/note="PCR was performed on DNA from Arabidopsis thaliana plants (T1) which were transformed with the T-DNA from vector PAC161. The lines contain one or more T-DNA insertions. The DNA fragment(s) resulting from the PCR were directly sequenced to determine the genomic sequence flanking the insertion. Sequences displaying significant similarity to the A. thaliana nuclear genome sequence were processed for submission. T-DNA derived sequences were removed"

BASE COUNT 168 a 58 c 70 g 146 t 31 others
ORIGIN

Query Match 4.6%; Score 146.4; DB 17; Length 473;
Best Local Similarity 90.8%; Pred. No. 1.2e-11;
Matches 167; Conservative 0; Mismatches 16; Indels 1; Gaps 1;

QY 1624 CTAGATGATGATATCCAAATACACTT-TGGATGTTTAACCTAATCTGTTCTC 1682
|||||
DB 186 CTAGATGATGATATCCAAATACACTTGGATGTTTAACTCTAATCTGTTCTT 127
QY 1683 CTACGCTATTAATATTAATCAATCGAGTAAAGTTTGTCTATTTTCGAGTCA 1742
|
DB 126 CCTAGTAAATATTAATCAATCGAGTAAAGTTTGTCTATTTTCGAGTCA 67
QY 1743 GAGGATTAACCTTAATGACTTTATTTTGAAGTAAAGTAACTTTTACTATGATTA 1802
|||||
DB 66 GAGGATTAACCTTAATGACTTTATTTTGAAGTAAAGTAACTTTTACTATGATTA 7
QY 1803 TTAC 1806
|||||
DB 6 TTAC 3

RESULT 4
CNS00EVL/C 1101 bp DNA linear GSS 04-JUN-1999
LOCUS CNS00EVL
DEFINITION Drosophila melanogaster genome survey sequence T7 end of BAC: BACR29B23 of RPI-98 library from Drosophila melanogaster (fruit fly), genomic survey sequence.
ACCESSION AL069706
VERSION AL069706.1 GI:4949849
KEYWORDS GSS.
SOURCE Drosophila melanogaster.
ORGANISM Drosophila melanogaster.
REFERENCE 1 (bases 1 to 1101)
AUTHORS Genoscope.
TITLE Direct Submission
JOURNAL Submitted (02-JUN-1999) Genoscope - Centre National de Sequenage : BP 191 91006 EVRY cedex - FRANCE (E-mail : sequef@genoscope.cns.fr - Web : www.genoscope.cns.fr)
COMMENT Determination of this BAC end sequence was carried out as part of a collaboration with the Berkeley Drosophila genome project (BDGP). The BDGP is constructing a physical map of the Drosophila

melanogaster genome using these BACs. For further information please see [http://www.fruitfly.org/TheBDGP/Drosophila melanogaster BAC library](http://www.fruitfly.org/TheBDGP/Drosophila_melanogaster_BAC_library) was prepared by Kazutoyo Oseguwa and Aaron Mammoser in Pieter de Jong's laboratory in the Department of Cancer Genetics at the Roswell Park Cancer Institute in Buffalo, NY. The library is named RPI-98 and was constructed by partial EcoRI digestion of Drosophila DNA provided by the BDGP from the isogenic strain y2; cn bw sp, the same strain used for the BDGP's p1 and EST libraries. A more detailed description of the library and how to order individual BAC clones, the entire library, or filters for hybridization from the BACPAC Resource Center can be found at http://bacpac.med.buffalo.edu/drosophila_bac.htm.

FEATURES
source
1. 1101
location/Qualifiers
/organism="Drosophila melanogaster"
/db_xref="taxon:7227"
/clone="BACR29B23"
/clone_lib="RPI-98"
/note="end : 17"

BASE COUNT 419 a 91 c 60 g 299 t 232 others
ORIGIN

Query Match 3.5%; Score 113.2; DB 17; Length 1101;
Best Local Similarity 35.8%; Pred. No. 4e-07;
Matches 232; Conservative 118; Mismatches 297; Indels 1; Gaps 1;

QY 1625 TAGTATGATGATATCCAAATACACTTGGATGTTTAACTTAATCTGTTCTTCC 1684
|
DB 1100 TTTTATTTTATTTTATTTTATTTTATTTTATTTTATTTTATTTTATTTTATTTTAT 1041
QY 1685 ACGGATTAATATTAATCAATCGAGTAAAGTTTGTCTATTTTCGAGTCAAGA 1744
|
DB 1040 TATATATATTTTAAATATATATTTTAAATATATTTTAAATATATTTTAAATATATA 981
QY 1745 AGGATTAACCTTAATGACTTTATTTTGAAGTAAAGTAACTTTTACTATGATTA 1804
|
DB 980 TATATATTTTAAATATATTTTAAATATATTTTAAATATATTTTAAATATATA 921
QY 1805 ACCGATGTTTGTGTCATTAATGAGCCCTCACTGATGATGATGATGATGATG 1864
|
DB 920 TTTTATTTTATTTTATTTTATTTTATTTTATTTTATTTTATTTTATTTTATTTTAT 861
QY 1865 CAATATTAATTAATGAATTCATGCTACTATCAATAGAGAACGCTGATTAAT 1924
|
DB 860 TTAATTTTATTTTATTTTATTTTATTTTATTTTATTTTATTTTATTTTATTTTAT 801
QY 1925 TTTAATTTAAGACAAATTTTGAAGTAAAGTAAATTTTCTAACAATTTTAAATA 1984
|
DB 800 TTTTATTTTATTTTATTTTATTTTATTTTATTTTATTTTATTTTATTTTATTTTAT 741
QY 1985 TGATGCTATTAATGATTTCCAT-TGCTTAAATATTTTATTTTATTTTATTTAGTTAA 2043
|
DB 740 TTTTATTTTATTTTATTTTATTTTATTTTATTTTATTTTATTTTATTTTATTTTAT 681
QY 2044 ATACATTTATGAACCAATATAGTGTGATTAATCAATATTCATTAATTTTGGAA 2103
|
DB 680 TTAATTTTATTTTATTTTATTTTATTTTATTTTATTTTATTTTATTTTATTTTAT 621
QY 2104 TCTACAAATTTAATTTTATTTAGTCAATTAATGATGATGATGATGATGATGATG 2163
|
DB 620 TTAATTTTATTTTATTTTATTTTATTTTATTTTATTTTATTTTATTTTATTTTAT 561
QY 2164 TTAACAGAACTTCAATTTTATTTTATTTTATTTTATTTTATTTTATTTTATTTTAT 2223
|
DB 560 TTTTATTTTATTTTATTTTATTTTATTTTATTTTATTTTATTTTATTTTATTTTAT 501
QY 2224 TTTTGTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 2271
|
DB 500 TTTTATTTTATTTTATTTTATTTTATTTTATTTTATTTTATTTTATTTTATTTTAT 453

RESULT 5
CNS00EVL

LOCUS	CNS00EVL	1101 bp	DNA	Linear	GSS 04-JUN-1998
DEFINITION	Drosophila melanogaster genome survey sequence T7 end of BAC: BACH29B23 of RPCT-98 library from Drosophila melanogaster (fruit fly), genomic survey sequence.				
ACCESSION	AL069706				
VERSION	AL069706.1	GI:4949849			
KEYWORDS	GSS.				
SOURCE	Drosophila melanogaster.				
ORGANISM	Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.				
REFERENCE	1 (bases 1 to 1101)				
AUTHORS	Genoscope.				
TITLE	Direct Submission				
JOURNAL	Submitted (02-JUN-1999) Genoscope - Centre National de Sequencage : BP 191 91006 EVRY cedex - FRANCE (E-mail : seqef@genoscope.cns.fr - Web : www.genoscope.cns.fr)				
COMMENT	determination of this BAC-end sequence was carried out as part of a collaboration with the Berkeley Drosophila Genome Project (BDGP). The BDGP is constructing a physical map of the Drosophila melanogaster genome using these BACs. For further information please see http://www.fruitfly.org The BDGP Drosophila melanogaster BAC library was prepared by Kazuo Oosawa and Aaron Mamosser at Pieter de Jong's laboratory in the Department of Cancer Genetics at the Roswell Park Cancer Institute in Buffalo, NY. The library is named RPCT-98 and was constructed by partial EcoRI digestion of Drosophila DNA provided by the BDGP from the isogenic strain y2; cn bw sp, the same strain used for the BDGP's p1 and Est libraries. A more detailed description of the library and how to order individual BAC clones, the entire library, or filters for hybridization from the BACPAC Resource Center can be found at http://bacpac.med.buffalo.edu/drosophila_bac.htm.				
FEATURES	Location/Qualifiers				
source	1..1101				
	/organism="Drosophila melanogaster"				
	/db_xref="taxon:7227"				
	/clone="BACR29B23"				
	/clone_lib="RPCT-98"				
	/note="end : 77"				
BASE COUNT	419 a 91 c 60 g 299 t 232 others				
ORIGIN					
Query Match	3.5%; Score 110.8; DB 17; Length 1101;				
Best Local Similarity	34.5%; Pred. No. 8.7e-07;				
Matches 223:	Conservative 128; Mismatches 290; Indels 5; Gaps 1;				
OY	1513 TAACCANTCGAACATTAATAATGCTTCTTTAAGCACTTAACATATTATGTGGTAG				
	::: ::: ::: ::: ::: ::: ::: ::: ::: ::: ::: :::				
DB	461 TTWTMMMMMMMAATVWTTWMAAAAMAAATTAATAATVAAAAMAMAMMTATTTTTMMWTTWA				
	::: ::: ::: ::: ::: ::: ::: ::: ::: ::: ::: :::				
OY	1573 TTGGAATTAATAAGACGCACTTGTAACATTTACAATTTAATATAGTACTAGTAGT				
	::: ::: ::: ::: ::: ::: ::: ::: ::: ::: ::: :::				
DB	521 TTWTTTTMMMTWTMTWTAATAAAAAAMMTAATTTAAAMWATAATTAAMATTAMAMW				
	::: ::: ::: ::: ::: ::: ::: ::: ::: ::: ::: :::				
OY	1633 GATATTCOCAATACARACTTGGATGTTTAACTTAATCTGTGTTCTCCACGGATA				
	::: ::: ::: ::: ::: ::: ::: ::: ::: ::: ::: :::				
DB	581 TATATTAATWATTAATTAATWATTAATTAATAAAAAATATTTTWTATATAAATTTTAAATA				
	::: ::: ::: ::: ::: ::: ::: ::: ::: ::: ::: :::				
OY	1693 AATATTAATCATCGACGTAATAAAAGTTTCTCTTATTTGCGATGCATGAACGATAA				
	::: ::: ::: ::: ::: ::: ::: ::: ::: ::: ::: :::				
DB	641 TTTAATTAATWATTAATTAATWATTAATWATTAATTAATTAATTAATTAATTAATTA				
	::: ::: ::: ::: ::: ::: ::: ::: ::: ::: ::: :::				
OY	1753 CCTAATGACTTAATTTTGGAAAATGTAACCTTTACTCATGATTAATTAACCGTAG				
	::: ::: ::: ::: ::: ::: ::: ::: ::: ::: ::: :::				
DB	701 AAAAAMATWMAATWMTWMTAATTAATTAATAAAMATAAAMAMATAAMWATATATWMA				
	::: ::: ::: ::: ::: ::: ::: ::: ::: ::: ::: :::				
OY	1813 TTTTGTGGCATAATGACGCCCTCTCAACATGTCGATAGTCATTTTCTGCAAATATT				
	::: ::: ::: ::: ::: ::: ::: ::: ::: ::: ::: :::				
DB	761 TATATTAATTTWMAATWMAATWMTATWMA-----TAAATATAMAMAMWATTAATWATA				
	::: ::: ::: ::: ::: ::: ::: ::: ::: ::: ::: :::				
OY	1873 AAATTAGAATTCAGTCTCATCATAGAGAAGAACGCTATTAATTAATTTT 1932				
	::: ::: ::: ::: ::: ::: ::: ::: ::: ::: ::: :::				

[illegible]

Db 1037 TTTTCTTTTNNNNAAATATTHMCACTTTTCATMCCTTCTATTTTCMTHTTTT 978
 1865 GAAATATTAATAGGAATTCATGCTACTATCATAGAGAAACGCGATATTACAT 1924
 Db 977 AAMATTAAAMATTATATTTCTTACHATTTTACACMCMATYACMCMACATTCACAT 918
 1925 TTTAATTTAAAGACAAATTTTGAAGAAATGTTAATTTCTAACAATTTTAAATA 1984
 Db 917 TTTAACTTAAGAAACATATTAATTAATTAATTAATTAATTAATTAATTAATTA 858
 1985 TGAAGCCATTAATGTAATTCCTATGCTCTTAAATATTTTAAATTAAGTATATA 2044
 Db 857 AAAAAAATTTTATTTTATTTTATTTTATTTTATTTTATTTTATTTTATTTA 798
 2045 TACATTAAGAACCAATTAATAGTGTGATTCATTAATCTCATTAATTTTTCGAAT 2104
 Db 797 AAAAAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 741
 2105 CTACAAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 2164
 Db 740 TTTAAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 681
 2165 TTAAGAAATCTTCAATTTTATTTTATTTTATTTTATTTTATTTTATTTTATTT 2224
 Db 680 TTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 621
 2225 TTTGTTGTCAGAGTATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 2281
 Db 620 TTTTWTWTAAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 564
 RESULT 7
 CNS01100/c 836 bp DNA linear GSS 26-JUL-1999
 LOCUS Drosophila melanogaster genome survey sequence T7 end of BAC
 DEFINITION BACN05N18 of DrosBAC library from Drosophila melanogaster (fruit
 fly), genomic survey sequence.
 ACCESSION AL069642
 VERSION AL069642.1 GI:5611253
 KEYWORDS GSS.
 SOURCE Drosophila melanogaster.
 ORGANISM Drosophila melanogaster.
 Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
 Neoptera; Endopterygota; Diptera; Brachyera; Muscomorpha;
 Ephydroidea; Drosophilidae; Drosophila.
 1 (bases 1 to 836)
 REFERENCE
 AUTHORS Direct Submission
 TITLE Submitted (23-JUL-1999) Genoscope - Centre National de Sequencage :
 JOURNAL BP 191 91006 EVRY cedex - FRANCE (E-mail : seqref@genoscope.cns.fr
 - Web : www.genoscope.cns.fr)
 COMMENT determination of this BAC-end sequence was carried out as part of a
 collaboration with the European Drosophila Genome Project (BDGP) -
 http://www.edgp.ebi.ac.uk -. This Drosophila melanogaster BAC
 library (Dros BAC) was made by Alain Billaud at CEPH (Centre
 d'Etude du Polymorphisme Humain) with funding provided by a MRC
 project grant. The DNA was prepared from embryos by Alain Bucheton
 and Genevieve Payan. It has been constructed in the vector
 pBelobAC11.
 FEATURES
 source 1. 836
 Location/Qualifiers
 /organism="Drosophila melanogaster"
 /db_xref="taxon:7227"
 /clone="BACN05N18"
 /clone_1lb="DrosBAC"
 /plasmid="pBelobAC11"
 /note="end : 77"
 BASE COUNT 379 a 58 c 42 g 269 t 88 others
 ORIGIN
 Query Match 3.1%; Score 98.4; DB 17; Length 836;
 Best Local Similarity 45.3%; Pred. No. 5.8e-05;

Matches 264; Conservative 24; Mismatches 290; Indels 5; Gaps 1;
 QY 1546 ATGCATCTTAAATATTTATTTAGTTGGAATTTAAGAGGACCTGATCATTA 1605
 Db 592 ATCTTNNNNANNN 533
 1606 CAATTTATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 1665
 Db 532 TTAATAATAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 473
 1666 CTAAATCTTGTCTCTCGCTACGCTAATTAATTAATTAATTAATTAATTAATTA 1725
 Db 472 TGAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 413
 1726 T-----TATTTGGGATGATGAGGATTAATTAATTAATTAATTAATTAATTA 1780
 Db 412 TTTAAATCTTTTATTAATTTTATTAATTAATTAATTAATTAATTAATTAATTA 353
 1781 AACCTTTTACATGATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 1840
 Db 352 TTTTAAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 293
 1841 AACTGTGATGATTAATTTTCTGCAATTAATTAATTAATTAATTAATTAATTAAT 1900
 Db 292 ATAAATTTTATTTTATTAATTTTATTAATTTTATTAATTTTATTAATTTTATTA 233
 1901 AGAAGAAACAGCTGATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 1960
 Db 232 AAATTAATTAATTTTATTTTATTTTATTTTATTTTATTTTATTTTATTTTATTT 173
 1961 ATTTCTAACAAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 2020
 Db 172 TTAATTAATTAATTTTATTAATTTTATTAATTTTATTAATTTTATTAATTTTATTA 113
 2021 TTTTATTAATTTTATTAATTTTATTAATTTTATTAATTTTATTAATTTTATTA 2080
 Db 112 AATTTAAATTAATTTTATTAATTTTATTAATTTTATTAATTTTATTAATTTTAA 53
 2081 ATCTCATTAATTAATTTTGAATTTGAAATTCACAAATTAATTAATTTT 2123
 Db 52 ATATTTTAAATAATTTTATTAATTTTATTAATTTTATTAATTTTATTAATTTAT 10
 RESULT 8
 CNS00E07/c 1101 bp DNA linear GSS 04-JUN-1999
 LOCUS Drosophila melanogaster genome survey sequence TET3 end of BAC:
 DEFINITION BACR29P01 of RPCI-98 library from Drosophila melanogaster (fruit
 fly), genomic survey sequence.
 ACCESSION AL069440
 VERSION AL069440.1 GI:4949583
 KEYWORDS GSS.
 SOURCE Drosophila melanogaster.
 ORGANISM Drosophila melanogaster.
 Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
 Neoptera; Endopterygota; Diptera; Brachyera; Muscomorpha;
 Ephydroidea; Drosophilidae; Drosophila.
 1 (bases 1 to 1101)
 REFERENCE
 AUTHORS Direct Submission
 TITLE Submitted (02-JUN-1999) Genoscope - Centre National de Sequencage :
 JOURNAL BP 191 91006 EVRY cedex - FRANCE (E-mail : seqref@genoscope.cns.fr
 - Web : www.genoscope.cns.fr)
 COMMENT determination of this BAC-end sequence was carried out as part of a
 collaboration with the Berkeley Drosophila Genome Project (BDGP).
 The BDGP is constructing a physical map of the Drosophila
 melanogaster genome using these BACs. For further information
 please see http://www.fruitfly.org The BDGP Drosophila
 melanogaster BAC library was prepared by Kazutoyo Oosagawa and
 Aaron Mammosser in Pieter de Jong's laboratory in the Department of
 Cancer Genetics at the Roswell Park Cancer Institute in Buffalo,
 NY. The library is named RPCI-98 and was constructed by partial
 EcoRI digestion of Drosophila DNA provided by the BDGP from the

isogenic strain y2; cn bw sp, the same strain used for the BDGP's P1 and EST libraries. A more detailed description of the library and how to order individual BAC clones, the entire library, or filters for hybridization from the BACPAC Resource Center can be found at http://bacpac.med.buffalo.edu/drosophila_bac.htm.

FEATURES

Location/Qualifiers

1..1101
/organism="Drosophila melanogaster"
/db_xref="taxon:7227"
/clone="BACR29P01"
/clone_1lb="RPCI-98"
/note="end: TERT3"
BASE COUNT 366 a 66 c 104 g 351 t 214 others
ORIGIN

Query Match 3.1%; Score 98.2; DB 17; Length 1101;
Best Local Similarity 41.0%; Pred. No. 5.3e-05;
Matches 229; Conservative 77; Mismatches 237; Indels 15; Gaps 2;

QY 1786 TTTTACTCATGATTAATTCACCGTATGTTTGTGCGCATATGACAGCCTCTACAACTG 1845
D 1031 TTTTMMAMAAACAMAAATTTTMTATTTTMTTAAATATTCATATTTTMAATACA 972
QY 1846 TGATAGCAATTTTTCGCAATATTAATAGCAATTCATGCTACTACATACAGAG 1905
D 971 WMAATTTTWTATCATATATTTTATATATACATTTTMAAATTAACWMAAT 912
QY 1906 AAACAGCTGAGTA-TTACATTTTAAATTAAGACAAATTTTGAATAATTTAAATTT 1964
D 911 TTAATAAACATTTTAAATTTTAAATTTTAAATTTTAAATTTTAAATTTTAAATTT 852
QY 1965 CTACATATTTATTAATATGATGCTATATGATTTCTGATGTTCTTAAATATTT 2024
D 851 AAAAATAAAATATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 792
QY 2025 TTTTATATTTAGTATTAATATACATTAATTAATTAATTAATTAATTAATTAAT 2084
D 791 WMTATTTTATTTTAAATTTTAAATTTTAAATTTTAAATTTTAAATTTTAAATTT 732
QY 2085 CCAATTAATTTTGAATCTACAAATTTTAAATTTTAAATTTTAAATTTTAAATTT 2144
D 731 TTTTATTTTAAATTAATTAATTTTAAATTTTAAATTTTAAATTTTAAATTTAA 672
QY 2145 AGTTCACAAAATTTTGTACAGCA-----AAGTTCACAAATTTTGT 2190
D 671 WTAATTTTAAATTAATTTTAAATTTTAAATTTTAAATTTTAAATTTTAAATTTT 612
QY 2191 TTTATGACAAAGAAATACAGATAGAAACTATTTTGTGGAATGAGATGATATA 2250
D 611 WMCCTTAAATTAATTAATTAATTTTAAATTTTAAATTTTAAATTTTAAATTTTAA 552
QY 2251 TACATTAAGCAATTTTAAATTTTAAATTTTAAATTTTAAATTTTAAATTTTAA 2310
D 551 ATTATATTAATTAATTTTAAATTTTAAATTTTAAATTTTAAATTTTAAATTTTAA 492
QY 2311 GTAGGTGTAATTAATTAAT 2328
D 491 TWAATTTTAAATTTTAAATTT 474

RESULT 9

AV71372/c

LOCUS AV71372 807 bp mRNA linear EST 16-OCT-2000

DEFINITION AV71372 DCB Homo sapiens cDNA clone DCBCH07 5', mRNA sequence.

ACCESSION AV71372

VERSION AV71372.1 GI:10814524

KEYWORDS

EST.

SOURCE

ORGANISM

REFERENCE

AUTHORS

FEATURES

Location/Qualifiers

1..807
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="DCBCH07"
/clone_1lb="DCB"
/cell_type="dendritic cells"
/dev_stage="mature"
/lab_host="BM25.8"
/note="Vector: pT1p1ex2; Site_1: sf11A; Site_2: sf11B"
BASE COUNT 388 a 50 c 20 g 345 t 4 others
ORIGIN

Query Match 3.1%; Score 98; DB 10; Length 807;
Best Local Similarity 46.2%; Pred. No. 6.7e-05;
Matches 323; Conservative 0; Mismatches 376; Indels 0; Gaps 0;

QY 1577 AATTATTAAGACGCACTTTACATTTCAATTTTATATTTAGTACTAGTGTGTT 1636
D 773 AATTTTGAATTAATTAATTTTAAATTTTAAATTAATTAATTAATTAATTTTAA 714
QY 1637 ATCCAAATACATCTTTGATGTTTAAATCTTATCTGTTCTCCACGCTAAATTA 1696
D 713 ATTATTAATTTTAAATTTTAAATTTTAAATTTTAAATTTTAAATTTTAAATTTTAA 654
QY 1697 TTAATTCAGAGTAAAAAGTTTGTCTTATTTTGGCATGATGAGATTAACCTTA 1756
D 653 AAAAATATATTAATTAATTTTAAATTTTAAATTTTAAATTTTAAATTTTAAATTT 594
QY 1757 ATGACTTAATTTTGAATTAATTAATTTTAAATTTTAAATTTTAAATTTTAAATTT 1816
D 593 TTAATTTTAAATTTTAAATTTTAAATTTTAAATTTTAAATTTTAAATTTTAAATTT 534
QY 1817 TGTTCATATATGACAGCTCTACACGCTGATGATGCTCAATTTTTCGCAATTAAT 1876
D 533 ATATATATTTATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 474
QY 1877 TAGCAATTCAGTCTACTATCAATAGAAAGAAAGCTGAGTATTTACATTTTAAAT 1936
D 473 TTGATTTTAGTGAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 414
QY 1937 ACAAAATTTTGAATTAATTAATTTTAAATTTTAAATTTTAAATTTTAAATTTTAA 1996
D 413 ATTAATTTTAAATTAATTAATTTTAAATTTTAAATTTTAAATTTTAAATTTTAAAT 354
QY 1997 TGTATTTCCATGTTCTTAAATTAATTTTAAATTTTAAATTTTAAATTTTAAATTT 2056
D 353 ATTAAGCAATTTTAAATTTTAAATTTTAAATTTTAAATTTTAAATTTTAAATTTTAA 294
QY 2057 CAATTAATGTTGGAATTTCAATATCTCATTTATTTTGAATTTCAATTAATTT 2116
D 293 CTTCCTTATTAATTAATTAATTTTAAATTTTAAATTTTAAATTTTAAATTTTAAAT 234
QY 2117 AATTTTATGCAATTAATTAATTTTAAATTTTAAATTTTAAATTTTAAATTTTAAAT 2176
D 233 AATTAAGCACTCTATAGAAAGAAATTAATTAATTAATTAATTAATTTTAAATTTTAA 174
QY 2177 CCAATTTTAAATTTTAAATTTTAAATTTTAAATTTTAAATTTTAAATTTTAAATTT 2236
D 173 GTTACTTTTAAATTAATTAATTTTAAATTTTAAATTTTAAATTTTAAATTTTAAATTT 114

QY 2237 TCGAAGTAGTATATACATTAGCAAAATTTTAAAAAT 2275
 || | | | | | | | | | | | | | | | | | | | |
 Db 113 TGAATATGAAAAAAATGAAATATATTTATTT 75

RESULT 10
 CNS020K7/c 1092 bp DNA linear GSS 12-MAY-2000
 LOCUS Tetradon nigroviridis genome survey sequence T7 end of clone
 DEFINITION 22211 of library G from Tetradon nigroviridis, genomic survey
 sequence.
 AL175696

ACCESSION
 VERSION AL175696.1 GI:7813753
 KEYWORDS GSS: genome survey sequence.
 SOURCE Tetradon nigroviridis.
 ORGANISM Tetradon nigroviridis.
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Actinopterygii; Neopterygii; Teleostei; Euteleostei; Neoteleostei;
 Acanthomorphi; Acanthopterygii; Percomorphi; Tetraodontiformes;
 Tetraodontidae; Tetraodon.
 1 (bases 1 to 1092)
 Roest-Crolius,H., Jallion,O., Dasilva,C., Bouneau,L., Fisher,C.,
 Bernot,A., Fizames,C., Whicker,P., Brottier,P., Quetier,F.,
 Saurin,W. and Weissenbach,J.
 Human gene number estimate provided by genome wide analysis using
 Tetradon nigroviridis DNA sequence
 Unpublished
 2 (bases 1 to 1092)
 Roest-Crolius,H., Jallion,O., Dasilva,C., Fizames,C., Fisher,C.,
 Bouneau,L., Billault,A., Quetier,F., Saurin,W., Bernot,A. and
 Weissenbach,J.
 Characterization and repeat analysis of the compact genome of the
 freshwater pufferfish Tetradon nigroviridis
 Unpublished
 3 (bases 1 to 1092)
 Genoscope.
 Direct Submission
 Submitted (12-APR-2000)
 This sequence is a single read and was generated as part of a large
 scale clone-end sequencing project of the Tetradon nigroviridis
 genome. For more information, please take a look at
<http://www.genoscope.cns.fr/tetraodon>.

FEATURES
 source
 1..1092
 /organism="Tetradon nigroviridis"
 /db_xref="taxon:99883"
 /clone="22211"
 /clone_1lb="g"
 /note="Genoscope sequence ID : COAG222CF06LPI-end : T7"

BASE COUNT 383 a 169 c 165 g 262 t 113 others

Query Match 3.1%; Score 98; DB 17; Length 1092;
 Best Local Similarity 43.3%; Pred. No. 5.7e-05;
 Matches 176; Conservative 58; Mismatches 170; Indels 2; Gaps 1;

QY 1762 TTGAATTTTGAATGTAACCTTTCTCATGATTAATACCGTATGTTTGTG 1821
 ||| | | | | | | | | | | | | | | | | | | |
 Db 1046 TTTTCTTTTAAATTAATTAATTTTAAATTTTAAATTTTAAATTTTAAATTTT 987

QY 1822 CCATAATGACAGCTCTACAGCTGATGCAATTTTTCGCAAAATTTAAATTAAGA 1881
 || | | | | | | | | | | | | | | | | | | | |
 Db 986 AAAATTAATTAATTAATTTTAAATTTTAAATTTTAAATTTTAAATTTTAA 927

QY 1882 ATTCATGCTACTATCATATGAAGAAGAGTGAATTTTAAATTTTAAAGACAA 1941
 || | | | | | | | | | | | | | | | | | | | |
 Db 926 AAAAATTTTAAATTAATTAATTTTAAATTTTAAATTTTAAATTTTAAATTT 867

QY 1942 ATTTTGAAGAAATGTT-ATAATTCCTAACAATATTAATTAATATGATGCTTAATG 1999
 || | | | | | | | | | | | | | | | | | | | |
 Db 866 ATTATTAATTAATTTTAAATTAATTTTAAATTTTAAATTTTAAATTTTAA 807

QY 2000 ATTCTCATGCTCTTAAATTAATTTTAAATTTAGTATATAATACATTAGCAACCA 2059
 || | | | | | | | | | | | | | | | | | | | |
 Db 806 ATTAATTAATTAATTTTAAATTAATTTTAAATTTTAAATTTTAAATTTTAA 747

QY 2060 TAATAGTGGGATTAATTAATTCATATATTTTGAATTCAGCAATTAATTAAT 2119
 || | | | | | | | | | | | | | | | | | | | |
 Db 746 TTTTCTTTTAAATTTTAAATTTTAAATTTTAAATTTTAAATTTTAAATTTT 687

QY 2120 ATTTAGTCAATACATGATAGCAAGCTCCAAAATTTTGT 2165
 || | | | | | | | | | | | | | | | | | | | |
 Db 686 ATTTTCTTTTAAATTTTAAATTTTAAATTTTAAATTTTAAATTTTAA 641

RESULT 11
 CNS003BD 1101 bp DNA linear GSS 03-JUN-1999
 LOCUS Drosophila melanogaster genome survey sequence TET3 end of BAC #
 DEFINITION BACR0808 of RPCI-98 library from Drosophila melanogaster (fruit
 fly) genomic survey sequence.
 AL064091
 VERSION AL064091.1 GI:4941847
 KEYWORDS GSS.
 SOURCE Drosophila melanogaster.
 ORGANISM Drosophila melanogaster.
 Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
 Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
 Phnydroidea; Drosophilidae; Drosophila.
 1 (bases 1 to 1101)
 Genoscope.
 Direct Submission
 Submitted (02-JUN-1999) Genoscope - Centre National de Sequencage :
 BP 191 91006 EVRY cedex - FRANCE (E-mail : sequef@genoscope.cns.fr)
 Web : www.genoscope.cns.fr
 Determination of this BAC-end sequence was carried out as part of a
 collaboration with the Berkeley Drosophila Genome Project (BDGP).
 The BDGP is constructing a physical map of the Drosophila
 melanogaster genome using these BACs. For further information
 please see <http://www.fruitfly.org> The BDGP Drosophila
 melanogaster BAC library was prepared by Kazutoyo Osoegawa and
 Aaron Krumholz in Pletier de Jong's laboratory in the Department of
 Cancer Genetics at the Roswell Park Cancer Institute in Buffalo,
 NY. The library is named RPCI-98 and was constructed by partial
 EcoRI digestion of Drosophila DNA provided by the BDGP from the
 isogenic strain y2; cn bw sp, the same strain used for the BDGP's
 pl and EST libraries. A more detailed description of the library
 and how to order individual BAC clones, the entire library, or
 filters for hybridization from the BACPAC Resource Center can be
 found at http://bacpac.med.buffalo.edu/drosophila_bac.htm.

FEATURES
 source
 1..1101
 /organism="Drosophila melanogaster"
 /db_xref="taxon:7227"
 /clone="BACR0808"
 /clone_1lb="RPCI-98"
 /note="end : TET3"

BASE COUNT 395 a 120 c 103 g 334 t 149 others

Query Match 2.9%; Score 94; DB 17; Length 1101;
 Best Local Similarity 37.0%; Pred. No. 0.00021;
 Matches 248; Conservative 80; Mismatches 342; Indels 0; Gaps 0;

QY 1640 CCAATATACATCTTGGATGTTTAACTTAATCTGTTCTCCAGCTAATTAATA 1699
 || | | | | | | | | | | | | | | | | | | | |
 Db 432 CGATTTCTTTAATTAACACACACCAACCTTTTCGCMCMCMCMCMCMCMCM 491

QY 1700 ATCATCGAGTAAATTAATTTTCTTATTTTGCATGATGATTAACCTAATG 1759
 || | | | | | | | | | | | | | | | | | | | |
 Db 492 AATTTTAAATTAATTAATTTTAAATTTTAAATTTTAAATTTTAAATTTT 551

QY 1760 ACTTAAATTTTGAATATGTAACCTTTTACTATATGATTAATTAACCGATGTTTGT 1819
 || | | | | | | | | | | | | | | | | | | | |
 Db 552 WATTTTATTAATTAATTAATTTTAAATTTTAAATTTTAAATTTTAAATTTT 611

tarabfontaww nrt:fgcc

900 AATGTTTWTAAATTWGTAACTTTTAAATTTGTAATAAAAAANNAATATGTA 959

QY 2400 TTGAGTAAATGCTGTTCTGACTATTGAGGGCAAAAAAGACAAATGCCAAAAGTCCTA 2459
 Db 960 TAAWAAATRTWMTGTATTAKAAATATWGRATATTAKAAAAAAATATAAAWARRKKT 1019
 QY 2460 CGGGTTTGA 2468
 Db 1020 GTRWTKWA 1028

RESULT 15
 CENS00B01/c 1101 bp DNA linear GSS 04-JUN-1999
 LOCUS
 DEFINITION Drosophila melanogaster genome survey sequence 17 end of BAC #
 BACR24D09 of RPCI-98 library from Drosophila melanogaster (fruit
 fly), genomic survey sequence.
 AL057419
 AL057419.1 GI:4937865

ACCESSION
 VERSION
 KEYWORDS
 SOURCE
 ORGANISM
 Drosophila melanogaster.
 Drosophila melanogaster
 Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
 Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
 Ephyroidae; Drosophilidae; Drosophila.
 1 (bases 1 to 1101)

REFERENCE
 AUTHORS
 TITLE
 JOURNAL
 Genoscope.
 Direct Submission
 Submitted (02-JUN-1999) Genoscope - Centre National de Sequencage :
 BP 191 91006 EVRY cedex - FRANCE (E-mail : seqref@genoscope.cns.fr
 - Web : www.genoscope.cns.fr)

COMMENT
 Determination of this BAC-end sequence was carried out as part of a
 collaboration with the Berkeley Drosophila Genome Project (BDGP).
 The BDGP is constructing a physical map of the Drosophila
 melanogaster genome using these BACs. For further information
 please see <http://www.fruitfly.org> The BDGP Drosophila
 melanogaster BAC library was prepared by Kazutoyo Osoegawa and
 Aaron Mamoser in Pieter de Jong's laboratory in the Department of
 Cancer Genetics at the Roswell Park Cancer Institute in Buffalo,
 NY. The library is named RPCI-98 and was constructed by partial
 EcoRI digestion of Drosophila DNA provided by the BDGP from the
 isogenic strain y2; cn bw sp, the same strain used for the BDGP's
 P1 and EST libraries. A more detailed description of the library
 and how to order individual BAC clones, the entire library, or
 filters for hybridization from the BACPAC Resource Center can be
 found at http://bacpac.med.buffalo.edu/drosophila_bac.htm.
 location/Qualifiers
 1. 1101
 /organism="Drosophila melanogaster"
 /db_xref="taxon:7227"
 /clone="BACR24D09"
 /clone_lib="RPCI-98"
 /note="end : 17"

BASE COUNT 420 a 56 c 48 g 261 t 316 others
 ORIGIN

Query Match 2.98; Score 93; DB 17; Length 1101;
 Best Local Similarity 39.2%; Pred. No. 0.00029;
 Matches 235; Conservative 78; Mismatches 278; Indels 8; Gaps 1;

QY 1809 TATGTTTGTGTCCTAATGACAGCCTGACAGCTGATACATATTTTCTGCAAA 1868
 Db 1096 TTWMTTATWMTATWMAAAATTAATTAATWTAATTTTATTTTAAATTTTAAATTTTAA 1037
 QY 1869 TATTAAATAGCAATTCATGCTACTATCAATAGAAAGAACACGCTGATTAATTTTA 1928
 Db 1036 TWAATTAATATWATTTTAAAAAAAATWMAATTTTAAATTTTAAATTTTAAATTTT 977
 QY 1929 ATTAAAGCAAAATTTTGAAGAAATGTTATTAATTTCTAACAATATTATTAATAATGAT 1988
 Db 976 TWAATTTTAAATTTTAAATTTTAAATTTTAAATTTTAAATTTTAAATTTTAAATTTT 917
 QY 1989 GCCTAATAGTATTTCTGCTGCTAAATATTTTATTTTATTTAGTTAATTAACA 2048

Db 916 ATAAATAATTTTAAATATTTTAAATATTTTAAATATTTTAAATATTTTAAATATTTTAA 857
 QY 2049 TTATGAACATATATAGTGTGTAATTCATATATCTCCATTAATTTTGAATCTAC 2108
 Db 856 TTAAATTTATAMAMT-----WATATTTTATTTTATTTTATTTTATTTTATTTTATTTT 805
 QY 2109 AAATTTATTAATTTTATGCTAATTAACAATGCTAGAAAGTCCAAAAATTTGTTAAC 2168
 Db 804 WAATWATTTTATTTTATTTTAAATTTTAAATTTTAAATTTTAAATTTTAAATTTTAAATTTT 745
 QY 2169 AGAACTCCAAATTTTATTTTATGAAACAAGAAATTAACAGATGAAGAACTATTTTG 2228
 Db 744 AAATATTAATAATTTTATTTTATTTTATTTTATTTTATTTTATTTTATTTTATTTTATTT 685
 QY 2229 TTGTAATGGAAGTATGATTAATATACATTAAGCAATTTTAAATTTATATTAACCTATA 2288
 Db 684 WTSTPAAATSVCASSBSVWTTATATTTTASAAAMWAKAMATTTTSTGSTRTTTATTT 625
 QY 2289 CGGCTCAAGATATGCTATGCTGTAATTAATTAATGATGCGCATTCAGATT 2348
 Db 624 TTASTSSATTAATTAATTTATTTATTTATTTATTTATTTATTTATTTATTTATTTATTTATTT 565
 QY 2349 GGGCAACATTAAGAAAGCAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 2407
 Db 564 ATTAATAARSATTAATAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 506

Search completed: June 11, 2003, 06:58:37
 Job time : 4856.18 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: June 10, 2003, 22:00:04 ; Search time 2933.8 seconds
(without alignments)
10921.753 Million cell updates/sec

Title: US-09-502-426b-1_COPY_2102_3202

Perfect score: 1101

Sequence: 1 aatcacaattattaatat.....gagagagagaactagctcc 1101

Scoring table: IDENTITY NUC

Gapop 10.0 , Gapext 1.0

Searched: 2054640 seqs, 14551402878 residues

Total number of hits satisfying chosen parameters: 4109280

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

GenEmbl:*

1: gb.ba.*

2: gb.htg.*

3: gb.in.*

4: gb.om.*

5: gb.ov.*

6: gb.pat.*

7: gb.ph.*

8: gb.pl.*

9: gb.pr.*

10: gb.ro.*

11: gb.sts.*

12: gb.sy.*

13: gb.un.*

14: gb.vi.*

15: em.ba.*

16: em.fun.*

17: em.hum.*

18: em.in.*

19: em.mu.*

20: em.om.*

21: em.or.*

22: em.ov.*

23: em.pat.*

24: em.ph.*

25: em.pl.*

26: em.ro.*

27: em.sts.*

28: em.un.*

29: em.vi.*

30: em.htg.hum.*

31: em.htg.inv.*

32: em.htg.other.*

33: em.htg.mus.*

34: em.htg.pln.*

35: em.htg.fod.*

36: em.htg.mam.*

37: em.htg.vrt.*

38: em.sy.*

39: em.htgo.hum.*

40: em.htgo.mus.*

41: em.htgo.other.*

Pred. No. is the number of results predicted by chance to have a

score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	1101	100.0	4818	8 AF044216	AF044216 Arabidops
2	1090	99.0	84196	8 AF044216	AL132979 Arabidops
3	62.6	5.7	4629	3 AF273674	AF273674 Plasmidiu
4	61.4	5.6	12029	3 AE001372	AE001372 Plasmidiu
5	60.8	5.5	129360	2 AC117079	AC117079 Dictyoste
6	60.6	5.5	3364	14 TSP418778	AJ418778 Tomato sp
7	60.6	5.5	171317	9 AC020941	AC020941 Homo sapi
8	60	5.4	233269	2 AC094291	AC094291 Rattus no
9	59.4	5.4	224635	2 AL732314	AL732314 Homo sapi
10	59	5.4	253305	3 PF043377	AL034559 Plasmidiu
11	58.6	5.3	163669	2 AC114238	AC114238 Rattus no
12	58.6	5.3	192389	9 AC007182	AC007182 Homo sapi
13	58	5.3	115489	2 AC117072	AC117072 Dictyoste
14	57.8	5.2	27291	2 AC115575	AC115575 Dictyoste
15	57.8	5.2	169546	2 AC004157	AC004157 Plasmidiu
16	57.6	5.2	35077	3 CEY3888A	AL021483 Caenorhab
17	57.6	5.2	144784	9 AC009224	AC009224 Homo sapi
18	57.6	5.2	188993	2 AC109656	AC109656 Rattus no
19	57.2	5.2	18701	2 AC116985	AC116985 Dictyoste
20	57.2	5.2	73666	9 HSJ333B15	AL109954 Human DNA
21	57	5.2	43666	3 CEY51H1A	AL032644 Caenorhab
22	57	5.2	104992	2 AC005504	AC005504 Plasmidiu
23	57	5.2	156060	2 AC004153	AC004153 Plasmidiu
24	56.8	5.2	48811	2 AC116550	AC116550 Dictyoste
25	56.8	5.2	160759	2 AC117082	AC117082 Dictyoste
26	56.6	5.1	6107	6 AX252043	AX252043 Sequence
27	56.6	5.1	6107	6 AX344427	AX344427 Sequence
28	56.6	5.1	6107	6 AX348822	AX348822 Sequence
29	56.6	5.1	56099	2 AC115598	AC115598 Dictyoste
30	56.6	5.1	67970	3 PF043377	AL031746 Plasmidiu
31	56.6	5.1	86515	2 AC127730	AC127730 Rattus no
32	56.6	5.1	159941	2 AC016220	AC016220 Homo sapi
33	56.6	5.1	178642	9 AC023122	AC023122 Homo sapi
34	56.4	5.1	65691	3 PF043377	297348 Plasmidiu
35	56.4	5.1	99003	2 AL390756	AL390756 Homo sapi
36	56.4	5.1	110000	2 PF043377	Continuation (3 of
37	56.4	5.1	124041	9 CNS01DSJ	AL121808 Human chr
38	56.4	5.1	183353	9 AC087071	AC087071 Homo sapi
39	56.4	5.1	185273	9 AC073320	AC073320 Homo sapi
40	56.2	5.1	45296	2 AC115613	AC115613 Dictyoste
41	56.2	5.1	163443	2 AC006280	AC006280 Plasmidiu
42	56.2	5.1	196149	2 AC004709	AC004709 Plasmidiu
43	56	5.1	204439	9 AF000807	AF000807 Homo sapi
44	55.8	5.1	2166	3 AF200327	AF200327 Plasmidiu
45	55.8	5.1	14867	3 AE001398	AE001398 Plasmidiu

ALIGNMENTS

RESULT 1
AF044216

LOCUS

DEFINITION

Arabisidopsis thaliana steroid 22-alpha-hydroxylase (DMF4) gene,
complete cds.

ACCESSION
AF044216

VERSION
AF044216.1 GI:2935341

KEYWORDS

SOURCE

ORGANISM

Arabisidopsis thaliana.

Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;

Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;

Rosidae; eurosids II; Brassicales; Brassicaceae; Arabidopsis.

1 (bases 1 to 4818)

Choe,S., Dilkes,B.P., Fujioka,S., Takatsuto,S., Sakurai,A. and

4818 bp DNA linear PLN 25-JUN-2001
Arabisidopsis thaliana steroid 22-alpha-hydroxylase (DMF4) gene,
complete cds.

Feldmann, K.A.
The DWF4 gene of Arabidopsis encodes a cytochrome P450 that mediates multiple 22alpha-hydroxylation steps in brassinosteroid biosynthesis
Plant Cell 10 (2), 231-243 (1998)
98158690
9490746
2 (bases 1 to 4818)
Choe, S., Dilkes, B.P., Aspiroz, R. and Feldmann, K.A.
Direct Submission
Submitted (22-JAN-1998) Plant Sciences, University of Arizona, Tucson, AZ 85721, USA
Location/Qualifiers
1. 4818
/organism="Arabidopsis thaliana"
/strain="Ws-2"
/db_xref="taxon:3702"
/chromosome="III"
/map="8 cm from nga6"
990. 4398
/gene="DWF4"
/note="dwarf4"
/allele="dwf4-1; dwf4-2; dwf4-3; dwf4-4"
990. 1055
/gene="DWF4"
Join(<1133. 1353, 1434. 1758, 1844. 1996, 2095. 2409, 2562. 2654, 2746. 2824, 2931. 3040, 3795. 4398)
/gene="DWF4"
/product="steroid 22-alpha-hydroxylase"
Join(1133. 1353, 1434. 1758, 1844. 1996, 2095. 2409, 2562. 2654, 2746. 2824, 2931. 3040, 3795. 4040)
/gene="DWF4"
/function="brassinosteroid biosynthetic enzyme"
/note="member of the cytochrome P450 superfamily; CYP90B1"
/codon_start=1
/product="steroid 22-alpha-hydroxylase"
/protein_id="AAC05093.1"
/db_xref="GI:2935342"
/translation="MFETEHHLLPLLLPSLLSLLFLILKRRNRKTRFLNPPGKS
GWPLFGETIGYLPYATATGLGFMQHVSKYKIVRSNLFGEPTIVSADAGLNRFILQ
NEGRLEFCSYPSIGSILGKMSLVGMHRDMSISINLSHARLRTILLKDVVERH
TLFVLDNQNSIFSRAQDEAKKFTENMAKHIMSDPGEETEQLKKEYVFMKGVVS
APLNQPTGATYKALSRATILIKFERKMEERKLDIKEEDDEEEVKTSDEAEMSKSDH
VRKQDDDLGLYKLSNLSTFEQILDILSLLFAGHETSSVAIALAIFLQACPKAV
EELREHLEIARAKELGESELNDDYKDMFTQCVINETLRLGNVWVFLHRKALKDV
RYKGYDIPSGWKLVPVISAHLNDSRYDQNLFPWRWQONNGASSSGSFSFTWGN
NMPFGGGPRLCAGSELAKLEMAVFIHLVLKFNWELAEDDQPFAPFVDFPNGLPIR
VSRIL"

BASE COUNT 1566 a 727 c 888 g 1637 t
ORIGIN

Query Match 100.0%; Score 1101; DB 8; Length 4818;
Best Local Similarity 100.0%; Pred. No. 4.3e-180;
Matches 1101; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 AATCTACAAATATTATTTAGTCAATAACATGCATAGAAAGTCCAAAAAATTT 60
DB 32 AATCTACAAATATTATTTAGTCAATAACATGCATAGAAAGTCCAAAAAATTT 91
QY 61 TGTAAACAGAACTTCCAAATTTTATTTATGGAACAAGAAATACAGATAGAAAC 120
DB 92 TGTAAACAGAACTTCCAAATTTTATTTATGGAACAAGAAATACAGATAGAAAC 151
QY 121 TATTTTGTGTGGAATGGAAGTAGTAATATACATTAAGCAAAATTTAAAAAATTTATATA 180
DB 152 TATTTTGTGTGGAATGGAAGTAGTAATATACATTAAGCAAAATTTAAAAAATTTATATA 211
QY 181 GCCTATACCGCTCAAAGTAGTTATCTAGTAGGTGTAATTAATATGATGGTCCGATT 240
DB 212 GCCTATACCGCTCAAAGTAGTTATCTAGTAGGTGTAATTAATATGATGGTCCGATT 271
QY 241 CAGAAATCGGACAAACATCAACCGAATTAATAATTAACCTTTAAATAATAATAAATTT 300
|||||

Db 272 CAGAAATGGGACAAACAATGAAACCGGAATTAATAATTAACCTTTAAAAATAAATAAATTT 331
QY 301 TGAGTAAATGTGTTTTCTGACTATTCTGAGGGGCAAAAAAGACAAATGCCAAAGATCTACG 360
Db 332 TGAGTAAATGTGTTTTCTGACTATTCTGAGGGGCAAAAAAGACAAATGCCAAAGATCTACG 391
QY 361 GGTTTGACTGTCCAGTTCGGTAATAATCTAATAACTCTGCTTTGACCGCACGCTCGTGT 420
Db 392 GGTTTGACTGTCCAGTTCGGTAATAATCTAATAACTCTGCTTTGACCGCACGCTCGTGT 451
QY 421 AGGGTCTCTGACATATTTTCACTGTTTACCCCTACTCGTGAGCCACCCCTTTTCCCAT 480
Db 452 AGGGTCTCTGACATATTTTCACTGTTTACCCCTACTCGTGAGCCACCCCTTTTCCCAT 511
QY 481 ATCCTAAGGTAATTTTGGAAATCCCAATTTAAACCGATTTGAGACCGTACCGGACTTCCT 540
Db 512 ATCCTAAGGTAATTTTGGAAATCCCAATTTAAACCGATTTGAGACCGTACCGGACTTCCT 571
QY 541 GGGATTCTGCTGGAGCATTTTATCAAAAAATTTATAGCACGAATGGGTTTATTATTTAAAA 600
Db 572 GGGATTCTGCTGGAGCATTTTATCAAAAAATTTATAGCACGAATGGGTTTATTATTTAAAA 631
QY 601 ACTCACAACTTGATCAGATAAAATTTTCAATAACACTTTTACGATGGATTCGTACGATCTA 660
Db 632 ACTCACAACTTGATCAGATAAAATTTTCAATAACACTTTTACGATGGATTCGTACGATCTA 691
QY 661 TCTAATGACTTTTTTTTTTCTACCGGTGGATGAAAGTTATAGTACTATATAGCCAGAGA 720
Db 692 TCTAATGACTTTTTTTTTTCTACCGGTGGATGAAAGTTATAGTACTATATAGCCAGAGA 751
QY 721 CAATTGATTATAGATATATCCATTAATCCATGATATTTATGATATAAATAGCTGTTAAAC 780
Db 752 CAATTGATTATAGATATATCCATTAATCCATGATATTTATGATATAAATAGCTGTTAAAC 811
QY 781 TATTTTCAGCATCGCAGCTTTCTGCAACTTTTGTGTTTAAATTAAGAGTTTAAATAATAA 840
Db 812 TATTTTCAGCATCGCAGCTTTCTGCAACTTTTGTGTTTAAATTAAGAGTTTAAATAATAA 871
QY 841 AGTATTAAGGAGCATACAGAGGCAACAAAAGTAAATGAACGAGAGAAACAAAAGCCAT 900
Db 872 AGTATTAAGGAGCATACAGAGGCAACAAAAGTAAATGAACGAGAGAAACAAAAGCCAT 931
QY 901 GAAGTCATGTTTGTAGTTTAAAGTAAATGAAGAGTATTTTAAATTTTAAATGACGATGA 960
Db 932 GAAGTCATGTTTGTAGTTTAAAGTAAATGAAGAGTATTTTAAATTTTAAATGACGATGA 991
QY 961 TAACAATATATTTCTGACTTTCTTAAACCCCTCTTACAAACAGAGCTCCCTTTT 1020
Db 992 TAACAATATATTTCTGACTTTCTTAAACCCCTCTTACAAACAGAGCTCCCTTTT 1051
QY 1021 CAGTAGAGTCCGATTTCCCAATCTTAAAGACAAAAGCCATTAGAAAGAGAAAGTGAGTGAG 1080
Db 1052 CAGTAGAGTCCGATTTCCCAATCTTAAAGACAAAAGCCATTAGAAAGAGAAAGTGAGTGAG 1111
QY 1081 AGAGAGAGAGAACTAGCTCC 1101
Db 1112 AGAGAGAGAGAACTAGCTCC 1132

RESULT 2
ATT3A5/c
LOCUS
DEFINITION
Arabidopsis thaliana DNA chromosome 3, BAC clone T3A5.
ACCESSION
AL132979
VERSION
AL132979.2 GI:6782244
KEYWORDS
SOURCE
Arabidopsis thaliana.
ORGANISM
Arabidopsis thaliana
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
Rosidae; eurosids II; Brassicales; Brassicaceae; Arabidopsiis.
1 (bases 1 to 84196)
REFERENCE
Bloembergen, H., Mewes, H.W., Lemcke, K., Mayer, K.F.X., Quettier, F. and


```

/number=6
complement(28374. .28698)
/number=7
complement(28699. .28778)
/number=7
complement(28779. .28999)
/number=8
38307. .41323
/genes="T3A5.50"
complement(join(38307. .38796,38877. .39090,39189. .39270,
40180. .40242,40337. .40401,40489. .40564,40655. .40734,
40805. .40919,41225. .41323))
/notes="Contains Eukaryotic putative RNA-binding region
RNP-1 signature AA179-186"
/codon_start=1
/product="U1 snRNP 70K protein"
/protein_id="CAB62436.1"
/db_xref="GI:6561970"
/translation="MGSDGDFLRNPNAVQARAKVQNRANVQLKMGQSHPTGLN
NLLKLFEPPLLEYKPPPEKRCPPYGMQAFVSFNAEPDPEYAPPPPELPSQKR
ERIKRLKEGVEKAEDLKRYDNPNDPNATGDPYKTLFVSLNYESSESKIKREFES
YGPVKRHLVDLTNPKGVAFIEYMTDMKAAV KQADGOKIDGRVLDVVERGT
VNPMPRLGGGLGTSRVGGEEIVGEOQPOGRTSQSEPSRPREERKSREKGERE
RSRELSHQPRERSDRPREDKHHRDQGGDRDRSDRDRDRDRDRDRDRDRDR
GDRTSRDRDRDRSRDKERDIYEGGEYEHGGGRSRDRAEYRGEPEETRGYYEDDQGD
TDRYSHRDKMEEDDFRYERYEYKRSRSREYVR"
complement(38307. .38796)
/number=1
complement(38797. .38876)
/number=1
complement(38877. .39090)
/number=2
complement(39091. .39188)
/number=2
complement(39189. .39270)
/number=3
complement(39271. .40179)
/number=3
complement(40180. .40242)
/number=4
complement(40243. .40336)
/number=4

```

Query Match 99.0%; Score 1090; DB 8; Length 84196;

Best Local Similarity 99.9%; Pred. No. 1.9e-178;

Matches 1101; Conservative 0; Mismatches 0; Indels 1; Gaps 1;

```

QY 1 AATCTACAAATATTAAATATTAGTCAATAACAATGCATAGAAAGTTCCAAAAAATTT 60
Db |
QY 61 TGTTAACAGAACTTCCAAATTTTATGGAACAAGAAATAACAGATAGAAAC 120
Db |
QY 30041 TGTTAACAGAACTTCCAAATTTTATGGAACAAGAAATAACAGATAGAAAC 29982
QY 131 TATTTTGTGTGGAATGAAGTAGTAATACATTAAGCAAAATTTAAAAATATATAA 180
Db |
QY 29981 TATTTTGTGTGGAATGAAGTAGTAATACATTAAGCAAAATTTAAAAATATATAA 29922
QY 181 GCCTATACGCGCTCAAACTGTTATCTAGTAGGTGTAATTAATATGCATGCGGATT 240
Db |
QY 29921 GCCTATACGCGCTCAAACTGTTATCTAGTAGGTGTAATTAATATGCATGCGGATT 29862
QY 241 CAGAAATGGGCAACAATGAAACGGAATTTAAATATTAACTTTAAATATAAATAAATTT 300
Db |
QY 29861 CAGAAATGGGCAACAATGAAACGGAATTTAAATATTAACTTTAAATATAAATAAATTT 29802
QY 301 TGAGTAAATGTGTTCTTCTGACTATTGAGGGGCAAAAAAGACAATGCCAAAGTCTACG 360
Db |
QY 29801 TGAGTAAATGTGTTCTTCTGACTATTGAGGGGCAAAAAAGACAATGCCAAAGTCTACG 29742
QY 361 GGTGTGACTGTCCAGTTCGGTAATTAATCTGTTCTTTGACCGCGCTCGTGT 420
Db |

```

```

Db 29741 GGTGTGACTGTCCAGTTCGGTAATTAATCTGTTGTGACCGCACGCTCGTGT 29682
QY 421 AGGGGTCTCTGACATTTTCACTGTGTACCCCTACTCGTGAGCCACCCCTTTTCCCAT 480
Db |
Db 29681 AGGGGTCTCTGACATTTTCACTGTGTACCCCTACTCGTGAGCCACCCCTTTTCCCAT 29622
QY 481 ATCTAAGGTAATTTTGGAAATCCCAATTTAAACCGATTGAGACCGTACCGGACTTCCT 540
Db |
Db 29621 ATCTAAGGTAATTTTGGAAATCCCAATTTAAACCGATTGAGACCGTACCGGACTTCCT 29562
QY 541 GGGATTCTCTGGAGCATTTATCAAAAATTTATAGCAGCAATGGGTTTATTAATTTAAA 600
Db |
Db 29561 GGGATTCTCTGGAGCATTTATCAAAAATTTATAGCAGCAATGGGTTTATTAATTTAAA 29502
QY 601 ACTCACAACTTGCATCAGATATAAATTTCAATAAACACTTTTACGATGATTCTGACGATCTA 660
Db |
Db 29501 ACTCACAACTTGCATCAGATATAAATTTCAATAAACACTTTTACGATGATTCTGACGATCTA 29442
QY 661 TCTAATGACTTTTTTTTTTCTACCGGTGGATGAAAGTTATAGTACTATTAGCCAGAGA 720
Db |
Db 29441 TCTAATGACTTTTTTTTTTCTACCGGTGGATGAAAGTTATAGTACTATTAGCCAGAGA 29382
QY 721 CAATTGATTATAGATATATCCATTATCCATGATATTATGATATAAATAGCTGTTAAAC 780
Db |
Db 29381 CAATTGATTATAGATATATCCATTATCCATGATATTATGATATAAATAGCTGTTAAAC 29322
QY 781 TATTTTCAGCATCGCAGCTTTCTGCAACTTTTGTGTTTAAATTTAAAGAGTTTAAATAATAA 840
Db |
Db 29321 TATTTTCAGCATCGCAGCTTTCTGCAACTTTTGTGTTTAAATTTAAAGAGTTTAAATAATAA 29262
QY 841 AGTATTAAAGAGGAGCATTAACGAGGCAACAAAAGTAAATGAACGAGGAGAAACAAAGCCAT 900
Db |
Db 29261 AGTATTAAAGAGGAGCATTAACGAGGCAACAAAAGTAAATGAACGAGGAGAAACAAAGCCAT 29202
QY 901 GAAGCTCATGCTGTTA-GTTTAAAGCTTAATAAGAGATTTTATTAATTTTAAATGACGATG 959
Db |
Db 29201 GAAGCTCATGCTGTTAGTTTAAAGCTTAATAAGAGATTTTATTAATTTTAAATGACGATG 29142
QY 960 ATAACAATTTATTTTCTGACTTCTTTTAAACCCCTCTTACAAACAGAGCTCCCTTTT 1019
Db |
Db 29141 ATAACAATTTATTTTCTGACTTCTTTTAAACCCCTCTTACAAACAGAGCTCCCTTTT 29082
QY 1020 TCAGTAGAAGTCCGATTTCCCAATCTTTAAAGACAAACCAATTTAGAAAGAGAAAGTGAAGTGA 1079
Db |
Db 29081 TCAGTAGAAGTCCGATTTCCCAATCTTTAAAGACAAACCAATTTAGAAAGAGAAAGTGAAGTGA 29022
QY 1080 GAGAGAGAGAGAACTAGCTCC 1101
Db |
Db 29021 GAGAGAGAGAGAACTAGCTCC 29000

```

RESULT 3

AF273674

LOCUS

DEFINITION

ACCESSION

VERSION

KEYWORDS

SOURCE

ORGANISM

REFERENCE

AUTHORS

TITLE

JOURNAL

MEDLINE

PUBMED

REFERENCE

AUTHORS

TITLE

JOURNAL

AF273674 4629 bp DNA linear INV 24-APR-2001
Plasmodium falciparum DNA-dependent RNA polymerase (rRNA) gene,
complete cds; nuclear gene for probable mitochondrial product.

AF273674

AF273674.1 GI:9857984

Plasmodium falciparum.

Plasmodium falciparum

Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.

Li J., Muga, J. A., Cermakian, N., Cedergren, R. and Feagin, J. E.

Identification and characterization of a Plasmodium falciparum RNA

polymerase gene with similarity to mitochondrial RNA polymerases

Mol. Biochem. Parasitol. 113 (2), 261-269 (2001)

21192559

11295180

2 (bases 1 to 4629)

Li J., Muga, J. A., Cermakian, N., Cedergren, R. and Feagin, J. E.

Direct Submission

Submitted (31-MAY-2000) Seattle Biomedical Research Institute, 4


```

Db      18550 AAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAATAATAATCTTGAATGGGATAATCT 18491
QY      723 ATT-GATTATAGATATATCCATTAATCCATGATATATTATGATATATAAATAGCTGTTAAACT 781
Db      18490 CTTCAATATAGTAATGACACTACTATAAAGTTTATATAATATATTATTGTTAACTACT 18431
QY      782 ATTTCAGCATCGCAGCTTCTCAGACTTTGTTTAAATTTAGAGCTTTAAATAAATAA 841
Db      18430 CTTTAAACAAACAAATAGAAAATAAACAACAAAAAATAAATAAATAAATAAATAA 18371
QY      842 GTATTAAAGAGCATAACGAGGACACAAAAAGTAATGAACGAGGAGAAACAAAGGCATG 901
Db      18370 AATAAAACCCATTAATTAATTTTAAAGAGATAAATATTATTGTTATTTAATAATAT 18311
QY      902 AAGCTCATTTGTTAGTTTAAAGCTTAAATAGACAGATTTTAAATTTT 949
Db      18310 TTATTTTATTTTATTTTATTTTATTTTATTTTATTTTATTTTATTTTATTTT 18263

RESULT 6
TSP418778/c
LOCUS      TSP418778
DEFINITION      3364 bp RNA linear VRL 02-MAY-2002
SOURCE      Tomato spotted wilt virus NSs gene and N gene, genomic RNA, isolate 10HK96.
ACCESSION      AJ418778
VERSION      1
KEYWORDS      N gene; N protein; NSs gene; NSs protein.
SOURCE      Tomato spotted wilt virus.
ORGANISM      Viruses; ssRNA negative-strand viruses; Bunyaviridae; Tospovirus.
REFERENCE      1
AUTHORS      Heinze,C., Willingmann,P. and Adam,G.
TITLE      Short intergenic regions of the S RNAs of Tomato spotted wilt tospovirus -not a species characteristic
JOURNAL      Unpublished
AUTHORS      Adam,G.
TITLE      2 (bases 1 to 3364)
JOURNAL      Direct Submission
FEATURES      source
               1..3364
               /organism="Tomato spotted wilt virus"
               /viral
               /isolate="10HK96"
               /specific_host="virginia tobacco"
               /db_xref="taxon:11613"
               /country="Bulgaria"
               /gene="NSs"
               /gene="NSs"
               /codon_start=1
               /product="NSs protein"
               /protein_id="CAD11445.1"
               /db_xref="GI:20429082"
               /translation="MSSSVYESIITQTRVWVGSTAGKAVVDSWIHELGTCSLLVQT
               QLYSDRSKSGFTYKAVGNLPCEEEILSRVYIPIEDDIDFSINIDDSVLAISVCS
               NTVNGVKHQHGLKVLSPDQLHSIGTSMNRSDIKRQLOEKDIPNDRYEIAANKG
               SLSCVKEHTYKIETCYNOALGVNVLSPNRNHWLYSKFSPNOVESNRTVNSLAV
               RSLTSAENNPNSQAFVKASTDSHPKLSLWLPKLVROVSTOKLFKAGDENKT
               FVLSTVCPNHSVETALNISVICKHQLPIKCKAPSELSWFSDLKEPYNIVHDPSY
               PORVHALLEHTHFAQVLCNNLQEDVIITLNNVELTPGKLDLGERLTYNSEDICR
               KYFSLKTECLPSNTQMSYLDISQIPSWKIDFARGGKIKISQPVSVAKSLKLDLSG
               IKKGSKISDGYASGSK"
               complement(2436..3212)
               /gene="N"
               /complement(2436..3212)
               /gene="N"
               /function="structural protein"
               /codon_start=1
               /product="N protein"
               /protein_id="CAD11446.1"

```

```

/db_xref="GI:20429083"
/translation="MSKVLTRESIVALLTQGDLEFEEDQNLVAFNFKTFCLNDLQ
IKRMSVISCLTFEPKNSIMKVIKQSDFTFGKITIKKTSRIGATQMTFRRLDSMIRV
RIVEETGNSENLNTIKSLASHPLIOAYGLPLDDAKSVRLAIMLGGSLPLIASVDSFE
MISVLAITYODANTKDIGDPKVDKTEALGKVCYTLVLSKAFEMNEDQVKKGREYAAI
LSSNPNAGSGINMGHYSETLKNFYEMFGVKKQAKLTELA"
BASE COUNT      1002 a 563 c 489 g 1310 t
ORIGIN
Query Match      5.5%; Score 60.6; DB 14; Length 3364;
Best Local Similarity 48.2%; Pred. No. 0.38;
Matches 171; Conservative 0; Mismatches 184; Indels 0; Gaps 0;
QY      1 AATCTACAATTAATTAATTTAGTCAATAACAATGCATAGAGAAGTTCCAAAAAATTT 60
Db      2100 AATAAAAAATAAAAAATAAAAAATAAAAAATAAAAAATAAAAAATAAAAAATA 2041
QY      61 TGTTAACAGAACTTCCAAATTTTTTTTTTTTATGGAACAGAAATAACAGATAGAAAAC 120
Db      2040 AAAAAATAAAAGTAAAAATAAGAAATAAAAAATAAAAAATAAAAAATAAAAAATA 1981
QY      121 TATTTTGTGTGGAATGGAATAGTATATACATTAAGCAAAATTTTAAAAAATATATATA 180
Db      1980 AATAAAAAATAAAAAATAAAAAATAAAAAATAAAAAATAAAAAATAAAAAATA 1921
QY      181 GCCTATACGCGCTCAAAAGTATGTTATCTAGTAGTGTCTAATTAATAATGCATGCGGATT 240
Db      1920 AATAAAAAATAAAAAATAAAAAATAAAAAATAAAAAATAAAAAATAAAAAATA 1861
QY      241 CAGATTGGGACACAAATGAAACGGAATTAATAATATTAACTTTAAAAATAATAAAAT 300
Db      1860 TAAAAATAAAAAATAAAAAATAAAAAATAAAAAATAAAAAATAAAAAATAATA 1801
QY      301 TGAGTAAATGTGTTTTCTGACTATTGAGGGGCAAAAAAGCAATGCCAAAGT 355
Db      1800 AATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAA 1746

RESULT 7
AC020941/c
LOCUS      AC020941
DEFINITION      Homo sapiens chromosome 5 clone CTD-231219, complete sequence.
ACCESSION      AC020941
VERSION      1
KEYWORDS      HTG.
SOURCE      Homo sapiens.
ORGANISM      Homo sapiens
               Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
               Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE      1 (bases 1 to 171317)
AUTHORS      DOE Joint Genome Institute and Stanford Human Genome Center.
TITLE      Direct Submission
JOURNAL      Unpublished
AUTHORS      DOE Joint Genome Institute.
TITLE      2 (bases 1 to 171317)
JOURNAL      DOE Joint Genome Institute.
AUTHORS      DOE Joint Genome Institute.
TITLE      Direct Submission
JOURNAL      Submitted (12-JAN-2000) Production Sequencing Facility, DOE Joint
               Genome Institute, 2800 Mitchell Drive, Walnut Creek, CA 94598, USA
REFERENCE      3 (bases 1 to 171317)
AUTHORS      DOE Joint Genome Institute and Stanford Human Genome Center.
TITLE      Direct Submission
JOURNAL      Submitted (01-JUN-2001) DOE Joint Genome Institute, 2800 Mitchell
               Drive, Walnut Creek, CA 94598, USA
COMMENT      On Jun 1, 2001 this sequence version replaced gi:12830142.
               Draft Sequence Produced by DOE Joint Genome Institute
               www.jgi.doe.gov
               Finishing Completed at Stanford Human Genome Center
               www-shgc.stanford.edu
               Quality: Phrap Quality >=40 99.7% of Sequence;
               Estimated Total Number of Errors is 2.7.
FEATURES      source
               1..171317
               /organism="Homo sapiens"

```



```
* 10040 10139: gap of unknown length
* 10140 11793: contig of 1654 bp in length
* 11794 11893: gap of unknown length
* 11894 13371: contig of 1478 bp in length
* 13372 13471: gap of unknown length
* 13472 14864: contig of 1393 bp in length
* 14865 14964: gap of unknown length
* 14965 16460: contig of 1496 bp in length
* 16461 16560: gap of unknown length
* 16561 17663: contig of 1103 bp in length
* 17664 17763: gap of unknown length
* 17764 19281: contig of 1518 bp in length
* 19282 19381: gap of unknown length
* 19382 21112: contig of 1731 bp in length
* 21113 21212: gap of unknown length
* 21213 22833: contig of 1621 bp in length
* 22834 22933: gap of unknown length
* 22934 24214: contig of 1281 bp in length
* 24215 24314: gap of unknown length
* 24315 25527: contig of 1213 bp in length
* 25528 25627: gap of unknown length
* 25628 27220: contig of 1593 bp in length
* 27221 27320: gap of unknown length
* 27321 28926: contig of 1606 bp in length
* 28927 29026: gap of unknown length
* 29027 30363: contig of 1337 bp in length
* 30364 30463: gap of unknown length
* 30464 31665: contig of 1202 bp in length
* 31666 31765: gap of unknown length
* 31766 33449: contig of 1584 bp in length
* 33450 33550: gap of unknown length
* 33551 36317: contig of 2868 bp in length
* 36318 36417: gap of unknown length
* 36418 37778: contig of 1361 bp in length
* 37779 37878: gap of unknown length
* 37879 38551: contig of 1973 bp in length
* 38552 39552: gap of unknown length
* 39552 42802: contig of 2851 bp in length
* 42803 42902: gap of unknown length
* 42903 45682: contig of 2780 bp in length
* 45683 45782: gap of unknown length
* 45783 49268: contig of 3486 bp in length
* 49269 49368: gap of unknown length
* 49369 51429: contig of 2060 bp in length
* 51429 51529: gap of unknown length
* 51529 54571: contig of 3043 bp in length
* 54572 54671: gap of unknown length
* 54672 57547: contig of 2875 bp in length
* 57547 57646: gap of unknown length
* 57647 60740: contig of 3094 bp in length
* 60741 60840: gap of unknown length
* 60841 63916: contig of 3076 bp in length
* 63917 64016: gap of unknown length
* 64017 67000: contig of 2984 bp in length
* 67001 67100: gap of unknown length
* 67101 70219: contig of 3119 bp in length
* 70220 70319: gap of unknown length
* 70320 72950: contig of 2631 bp in length
* 72951 73050: gap of unknown length
* 73051 76802: contig of 3752 bp in length
* 76803 76902: gap of unknown length
* 76903 79990: contig of 3088 bp in length
* 79991 80090: gap of unknown length
* 80091 82745: contig of 2655 bp in length
* 82746 82845: gap of unknown length
* 82846 87725: contig of 4880 bp in length
* 87726 87825: gap of unknown length
* 87826 91178: contig of 3353 bp in length
* 91179 91278: gap of unknown length
* 91279 94533: contig of 3255 bp in length
* 94534 94634: gap of unknown length
* 94634 95871: contig of 5238 bp in length
* 95872 99971: gap of unknown length
```

```
* 99972 103742: contig of 3771 bp in length
* 103743 103842: gap of unknown length
* 103843 108877: contig of 5035 bp in length
* 108878 108977: gap of unknown length
* 108978 114019: contig of 5042 bp in length
* 114020 114119: gap of unknown length
* 114120 119231: contig of 5112 bp in length
* 119232 119331: gap of unknown length
* 119332 125649: contig of 6318 bp in length
* 125650 125749: gap of unknown length
* 125750 132887: contig of 6938 bp in length
* 132888 132987: gap of unknown length
* 132988 139741: contig of 6954 bp in length
* 139742 139841: gap of unknown length
* 139842 146349: contig of 6508 bp in length
* 146350 146449: gap of unknown length
* 146450 155325: contig of 8876 bp in length
* 155326 155425: gap of unknown length

Query Match      5.4%; Score 60; DB 2; Length 233269;
Best Local Similarity 51.3%; Pred No. 0.2;
Matches 138; Conservative 0; Mismatches 131; Indels 0; Gaps 0;

QY 6 ACAAATTAATATATTAGTCAATAACAATGATAGATAGAGTCCAAAAAATTTGTGTA 65
Db 1391 AAAAAATTTTGTGTTTTTAAAAAGTTAAAAAATTTTAAAAAATTTTAAAAA 1450
QY 66 ACAGAACTTCCAAATTTTTTTTTTATGGAACGAAGTAACAGATAGAAAACTATTT 125
Db 1451 TTTTAAAAAATATTTTTTTTTTTTTTTGAAAAAATTTTAAAAAATTTTAAAAA 1510
QY 126 TGTGTGGAATGGAAGTAGTATATATACATTAAGCAATTTTAAAAAATTTATATAAGCCCTA 185
Db 1511 TTTTAAAAAATATAAAAAATTTTGTGTAATTTTAAAAAATTTTAAAAAATTTT 1570
QY 186 TACGGCTCAAGTAGTATCTAGTAGTGAATTAATAATGATGCGGATTCAGAA 245
Db 1571 TTTTAAAAAATATTTTTTTTTTTTTTTCAAAAAATTTTAAAAAATTTTAAAAA 1630
QY 246 TTGGGACCAATGAACGGAATTTAAAA 274
Db 1631 TTTAAAAAATAAAAAATAAAAA 1659
```

RESULT 9

AL732314/C AL732314 224635 bp DNA linear HTG 17-AUG-2002
LOCUS Homo sapiens chromosome X clone RP13-465B17, *** SEQUENCING IN
DEFINITION PROGRESS ***, 8 unordered pieces.

ACCESSION

AL732314

VERSION

AL732314.10 GI:22415930

KEYWORDS

HTG; HTGS_PHASE1; HTGS_ACTIVEFIN; HTGS_DRAFT; HTGS_FULLTOP.

SOURCE

human.

ORGANISM

Homo sapiens

REFERENCE

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

AUTHORS

1 (bases 1 to 224635)

TITLE

Whitehead, S.

JOURNAL

Direct Submission

Submitted (16-AUG-2002) Wellcome Trust Sanger Institute, Hinxton,

Cambridgeshire, CB10 1SA, UK. E-mail enquiries:

humquery@sanger.ac.uk Clone requests: clonerequest@sanger.ac.uk

On Aug 21, 2002 this sequence version replaced gi:22204483.

----- Genome Center

Center: Wellcome Trust Sanger Institute

Center code: SC

Web site: <http://www.sanger.ac.uk>Contact: humquery@sanger.ac.uk

----- Project Information

Center project name: b8465b17

----- Summary Statistics

Assembly program: XGAP4; version 4.5

Chemistry: Dye-terminator; 100% of reads

Consensus quality: 222703 bases at least Q40
 Consensus quality: 223253 bases at least Q30
 Consensus quality: 223562 bases at least Q20
 Insert size: 223935; sum-of-contigs
 Insert size: 183401; 20.9% error; agarose-fp
 Quality coverage: 6.54x in Q20 bases; sum-of-contigs Quality
 coverage: 8.94x in Q20 bases; agarose-fp

 * NOTE: This is a 'working draft' sequence. It currently
 * consists of 8 contigs. The true order of the pieces
 * is not known and their order in this sequence record is
 * arbitrary. Gaps between the contigs are represented as
 * runs of N, but the exact sizes of the gaps are unknown.
 * This record will be updated with the finished sequence
 * as soon as it is available and the accession number will
 * be preserved.

1 3852: contig of 3852 bp in length
 * 3853 3952: gap of 100 bp
 * 3953 36565: contig of 32613 bp in length
 * 36566 36665: gap of 100 bp
 * 36666 70753: contig of 34088 bp in length
 * 70754 70853: gap of 100 bp
 * 70854 141740: contig of 70887 bp in length
 * 141741 141840: gap of 100 bp
 * 141841 180105: contig of 38265 bp in length
 * 180106 180205: gap of 100 bp
 * 180206 187608: contig of 100 bp
 * 187609 187708: gap of 100 bp
 * 187709 191079: contig of 3371 bp in length
 * 191080 191179: gap of 100 bp
 * 191180 224635: contig of 33456 bp in length.

FEATURES

source

1. 224635
 /organism="Homo sapiens"
 /db_xref="taxon:9606"
 /chromosome="X"
 /clone_lib="RPCI-13.2"
 misc_feature
 1. 3852
 /note="assembly_fragment:02748
 fragment_chain:1"
 misc_feature
 3953..36565
 /note="assembly_fragment:04038
 fragment_chain:1"
 misc_feature
 36666..70753
 /note="assembly_fragment:01967
 fragment_chain:1"
 misc_feature
 70854..141740
 /note="assembly_fragment:03650
 fragment_chain:1"
 misc_feature
 141841..180105
 /note="assembly_fragment:03847
 fragment_chain:1"
 misc_feature
 180206..187608
 /note="assembly_fragment:01716"
 misc_feature
 187709..191079
 /note="assembly_fragment:02259"
 misc_feature
 191180..224635
 /note="assembly_fragment:04008"
 BASE COUNT 56417 a 52985 c 53054 g 61479 t 700 others
 ORIGIN

Query Match

Best Local Similarity 5.4%; Score 59.4; DB 2; Length 224635;
 Matches 165; Conservative 0; Mismatches 176; Indels 0; Gaps 0;

Qy 6 ACAATTAATTAATTTAGTCATACATGCAATGCAAGAGTTCACAAAAAATTTTGTA 65
 Db 197253 AGAATAATTAATTAATTTAGTCATACATGCAATGCAAGAGTTCACAAAAAATTTTGTA 65
 Qy 66 ACAGAACTTCCAAATTTTTTTTTTTTATGGACAGAAATACACATAGAAAACTATT 125
 Db 197193 ATTTAATAATAAATTAATTTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTT 197134

QY 126 TGTCTGGGAATGGAAGTAGTATATACATTAAGCAAAATTTAAAAAATTTATATAGCCTA 185
 Db 197133 TAATAAATATTTTATATATTAATTAATAACATATAAAATCTTTAAATACAAATATATA 197074
 QY 186 TAGCGGCTCAAAAGTAGTATCTAGTAGGTGTAATTAATTAATGATGCGATCGGATTCAGAA 245
 Db 197073 TATTTCTTAATATTTTAATATATATGTTATATATTTTAAATATTTTATTTATT 197014
 QY 246 TTGGGACACATCAATGAACGGGAATTAATTAATTAATTAATTAATTAATTAATTAATTTAGT 305
 Db 197013 ATATATTTA 196954
 QY 306 AAATGTGTTTCTGACTATTGAGGGCGCAAAAAAGACAAT 346
 Db 196953 ATATATTTAACATATTTCTGTTAATATATATATAAATTTAAAT 196913

RESULT 10
PFMAL3P7

LOCUS

DEFINITION

ACCESSION

VERSION

KEYWORDS

SOURCE

ORGANISM

REFERENCE

AUTHORS

TITLE

JOURNAL

MEDLINE

PUBMED

REFERENCE

AUTHORS

JOURNAL

AUTHORS

TITLE

JOURNAL

COMMENT

FEATURES

source

1. 253305

/organism="Plasmodium falciparum 3D7"

/strain="3D7"

/db_xref="taxon:36329"

Massey, E., Mawhiney, E., McLeod, M.P., Meador, M., Mei, G., Metzker, M.,
Miner, G., Ninar, Z., Mitchell, T., Mohabbat, K., Morgan, M., Morris, S.,
Mosser, M., Neal, D., Newton, J., Newton, N., Nguyen, A., Nguyen, N.,
Nguyen, N., Nickerson, E., Nwokenkwo, S., Oguh, M., Okwuonu, G.,
Oragunye, N., Ovielo, R., Pace, A., Payton, B., Peery, J., Perez, L.,
Peters, L., Pickens, R., Primus, E., Pu, L.L., Quiles, J. M., Ren, Y.,
Rives, M., Rojas, A., Rojokoban, I., Rolfe, M., Ruiz, S., Savery, G.,
Scherer, S., Scott, G., Shen, H., Shooshtari, N., Sisson, I.,
Sodergren, E., Sonathe, T., Sparks, A., Stanley, H., Stone, H.,
Sutton, A., Svatek, A., Taylor, P., Tamerisa, A., Tamerisa, K., Tang, H.,
Tansey, J., Taylor, C., Taylor, T., Telford, B., Thomas, N., Thomas, S.,
Usmani, K., Vasquez, L., Vera, V., Villalon, D., Vinson, R., Wang, Q.,
Wang, S., Ward-Moore, S., Warren, R., Washington, C., Watlington, S.,
Williams, G., Williamson, A., Wleczyk, R., Wooden, S., Worley, K.,
Wu, C., Wu, Y., Wu, Y.F., Zhou, J., Zorrilla, S., Nelson, D.,
Weinstock, G. and Gibbs, R.

Submitted (07-MAR-2002) Human Genome Sequencing Center, Department
of Molecular and Human Genetics, Baylor College of Medicine, One
Baylor Plaza, Houston, TX 77030, USA
3 (bases 1 to 165869)
Worley, K.C.
Direct Submission
Submitted (13-JUL-2002) Human Genome Sequencing Center, Department
of Molecular and Human Genetics, Baylor College of Medicine, One
Baylor Plaza, Houston, TX 77030, USA
On Jul 12, 2002 this sequence version replaced gi:20467600.

Center: Baylor College of Medicine
Center code: BCM
Web site: <http://www.hgsc.bcm.tmc.edu/>
Contact: hgsc-help@bcm.tmc.edu
----- Project Information -----
Center project name: GQXB
Center clone name: CH230-336N7
----- Summary Statistics -----

Chemistry: Dye-terminator Big Dye: 100% of reads
 Assembly program: Phrap: version 0.990329
 Consensus quality: 128160 bases at least Q40
 Consensus quality: 133217 bases at least Q30
 Consensus quality: 136246 bases at least Q20

```

* NOTE: Estimated insert size may differ from sequence length
* (see http://www.hgsc.bcm.tmc.edu/docs/genbank_draft_data.html).
* NOTE: This is a 'working draft' sequence. It currently
* consists of 49 configs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the configs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.

```

	1	1074: contig of 1074 bp in length
*	1075	1174: gap of unknown length
*	1175	2179: contig of 1005 bp in length
*	2180	2279: gap of unknown length
*	2280	3300: contig of 1021 bp in length
*	3301	3400: gap of unknown length
*	3401	4592: contig of 1192 bp in length
*	4593	4692: gap of unknown length
*	4693	6219: contig of 1527 bp in length
*	6220	6319: gap of unknown length
*	6320	7439: contig of 1120 bp in length
*	7440	7539: gap of unknown length
*	7540	9318: contig of 1779 bp in length
*	9319	9418: gap of unknown length
*	9419	10930: contig of 1512 bp in length
*	10931	11030: gap of unknown length

RESULT 11
 AC114238
 LOCUS
 DEFINITION
 AC114238
 AC114238 bp DNA linear HTG 13-JUL-2002
 *** 49 unordered pieces. *** SEQUENCING IN PROGRESS
 AC114238
 AC114238.6 GI:21738343
 HTG; HTGS_PHASE1.
 Norway rat.
 Rattus norvegicus
 Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;
 Rattus.
 1 (bases 1 to 165669)
 Muzny,D.M., Adams,C., Adio-Oduola,B., Ali-Osman,F.R., Allen,C.,
 Alsbrooks,S.L., Amaratunge,H.C., Are,J.R., Ayele,M., Banks,T.,
 Barbara,J., Benton,J., Bimage,K., Blankenburg,K., Bonnin,D.,
 Bouck,J., Bowie,S., Brieva,M., Brown,E., Brown,M., Bryant,N.P.,
 Buahy,C., Burch,P., Burkett,C., Burrell,K.L., Byrd,N.C.,
 Carrott,T.F., Carter,M., Cavazos,S.R., Chacko,J., Chavez,D.,
 Chen,G., Chen,R., Chen,Z., Chowdry,I., Christopoulos,C.,
 Cleveland,C.D., Cox,C., Coyle,M.D., Dathorne,S.R., David,R.,
 Delavela,M.L., Davis,C., Davy-Carroll,L., Dederich,D.A.,
 Delaney,K.R., Delgado,O., Denn,A.L., Ding,Y., Dinh,H.H.,
 Douthwaite,K.J., Draper,H., Dugan-Rocha,S., Durbin,K.J.,
 Earhart,C., Edgar,D., Edwards,C.C., Elhaj,C., Fails,T.,
 Ferraguto,D., Flagg,N., Ford,J., Foster,C., Escotto,M.,
 Gabisi,A., Gao,J., Garcia,A., Garner,T., Garza,N., Gill,R.,
 Gorrell,J.H., Guevara,W., Gunaratne,P., Hale,S., Hamilton,K.,
 Harris,C., Harris,K., Hart,M., Haviak,P., Hawes,A., Hernandez,J.,
 Hernandez,O., Hodgson,A., Hogues,M., Holloway,C., Hollins,B.,
 Honsi,F., Howard,S., Huber,J., Hulyk,S., Hume,J., Jackson,L.E.,
 Jacobson,B., Jia,Y., Johnson,R., Jolivet,S., Joudah,S.,
 Karlsson,E., Kelly,S., Khan,U., King,L., Korvah,J., Kovar,C.,
 Kratovic,J., Kureshi,A., Landry,N., Leal,B., Lewis,L.,
 Li,J., Li,Z., Lichtarge,O., Lieu,C., Liu,J., Liu,W., Loulseged,H.,
 Lozado,R.J., Lu,X., Lucier,A., Lucier,R., Luna,R., Ma,J.,
 Maheshwari,M., Mapua,P., Martin,R., Martindale,A., Martinez,E.,

[illegible]

Query Match	Score	DB	Length
Best Local Similarity	52.7%	DB 9	Length 192389;
Matches 127;	Conservative	0;	Mismatches 114; Indels 0; Gaps
QY	97	AACRAGAATAACAGATAGAAAACACTATTTTGTGTGGAATGGAAGTAGTAAATATACATTA	
Db	176639	AAATAGAGATAATAATATCCAGGCTATTTGTAGGGGCTCTGTTAGAGCTCAATATAGCATA	
QY	157	AGCAAAATTTAAAAAATTTATATAGGCTATACGGCTCAAGATGTTATCTAGTAGTGTG	
Db	176699	ATGATATGGAATTTACTTCTTAAGTCTATACAAATGTAAGCATATATTTTACTGTT	
QY	217	TAATTAATATGATGTCGGATTCAGAAATTTGGGCAACAACATGAAACGGAATTAATAATA	
Db	176759	TAATATTGAATTTATGTCGGCTCTCTGTAAGATACAAGATTTACTTTGTAATTTGAAAAA	
QY	277	TTAACTTTAAAAATAAATAAATTTGAGTAAATGTTTTCGACTATGAGGGGCAAAA	
Db	176819	ATTCTTGAACATTAGAAGCTTTGAGGAAACGTTCTTCTATCACTGTTAGGTCATAGAA	
QY	337	A 337	
Db	176879	A 176879	
RESULT 13			
AC117072/c			
LOCUS	AC117072	115489 bp	DNA linear HTG 18-MAY-2004
DEFINITION	Dictyostelium discoideum chromosome 2 map 3426827-3542314 strain AX4, *** SEQUENCING IN PROGRESS ***, in ordered pieces.		
ACCESSION	AC117072.1	GI:20066239	
VERSION	HTG; HTGS_PHASE2.		
KEYWORDS	Dictyostelium discoideum.		
SOURCE	Dictyostelium discoideum.		
ORGANISM	Eukaryota; Mycetozoa; Dictyosteliida; Dictyostelium.		
REFERENCE	1 (bases 1 to 115489)		
AUTHORS	Gloeckner, G., Eichinger, L., Szafranski, K., Pachebat, J., Dear, P., Lehmann, R., Baumgart, C., Parra, G., April, J. F., Guigo, R., Kumpf, H., Tunggal, B., Cox, E., Quail, M. A., Platzer, M., Rosenthal, A. and Noegel, P. A.		
TITLE	Sequence and Analysis of Chromosome 2 of Dictyostelium		
JOURNAL	The Unpublished		
REMARK	The Dictyostelium Genome Sequencing Consortium		
REFERENCE	2 (bases 1 to 115489)		
AUTHORS	Baumgart, C.		
TITLE	Direct Submission		
JOURNAL	Submitted (06-APR-2002)		
REFERENCE	3 (bases 1 to 115489)		
AUTHORS	Biotechnology, Beutenbergstr. 11, Jena 07745, Germany		
TITLE	Baumgart, C.		
JOURNAL	Submitted (18-MAY-2002)		
REFERENCE	Genome Analysis, Institute of Molecular		
AUTHORS	Biotechnology, Beutenbergstr. 11, Jena 07745, Germany		
TITLE	Direct Submission		
JOURNAL	Submitted (18-MAY-2002)		
COMMENT	Genome Analysis, Institute of Molecular		
	Biotechnology, Beutenbergstr. 11, Jena 07745, Germany		
	CDS predictions from Genid may contain errors. Further information is available from IMB Jena, Department of Genome Analysis		
	(http://genome.imb-jena.de/dictyostelium/)		
	and the University Cologne, Institute for Biochemistry I		
	(http://www.uni-koeln.de/dictyostelium/project.shtml)		
	Funding		
	Agency : Deutsche Forschungsgemeinschaft (DFG).		
	* NOTE: This is a 'working draft' sequence.		
	* This sequence will be replaced		

Search completed: June 11, 2003, 01:49:57
Job time : 2938.8 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: June 10, 2003, 21:59:33 ; Search time 252.042 Seconds
(without alignments)
9837.455 Million cell updates/sec

Title: US-09-502-426b-1_copy_2102_3202
Perfect score: 1101
Sequence: 1 aatatacaattattataat.....gagagagagaactagctcc 1101

Scoring table: IDENTITY_NUC
Gapop 10.0, Gapext 1.0

Searched: 2185239 seqs, 1125999159 residues

Total number of hits satisfying chosen parameters: 4370478

Minimum DB seq length: 0
Maximum DB seq length: 2000000000
Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : N_Geneseq_101002.*

```
1: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1980.DAT.*
2: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1981.DAT.*
3: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1982.DAT.*
4: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1983.DAT.*
5: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1984.DAT.*
6: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1985.DAT.*
7: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1986.DAT.*
8: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1987.DAT.*
9: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1988.DAT.*
10: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1989.DAT.*
11: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1990.DAT.*
12: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1991.DAT.*
13: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1992.DAT.*
14: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1993.DAT.*
15: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1994.DAT.*
16: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1995.DAT.*
17: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1996.DAT.*
18: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1997.DAT.*
19: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1998.DAT.*
20: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1999.DAT.*
21: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA2000.DAT.*
22: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA2001A.DAT.*
23: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA2001B.DAT.*
24: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA2002.DAT.*
```

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	1101	100.0	6888	21	AAA59599
2	56.6	5.1	6107	24	ABL70390
3	56.6	5.1	6107	24	AAS61342
4	56.6	5.1	6107	24	ABK31431
5	54.2	4.9	875	21	AAA01920
6	54.2	4.9	6113	24	ABL32803
7	53.8	4.9	37973	24	ABL34197
8	53.8	4.9	83391	24	ABQ67094
9	53.4	4.9	17131	24	ABL33053

	10	53.2	4.8	8170	24	ABK28257
	11	53	4.8	54108	24	ABK22782
	12	53	4.8	57273	24	ABK22784
	13	53	4.8	66933	22	ABAB2625
	14	53	4.8	72049	22	ABAB2623
c	15	52.6	4.8	16228	24	ABL70459
c	16	52.6	4.8	16228	24	AAS61424
c	17	52.4	4.8	4501	24	ABK33968
c	18	52.4	4.8	5940	21	AAA70105
c	19	52.4	4.7	7746	24	ABK40047
c	20	52.2	4.7	7746	24	ABL33856
c	21	52.2	4.7	78925	21	ABK38888
c	22	52	4.7	5309	22	AAS46527
c	23	52	4.7	5309	24	ABK40039
c	24	52	4.7	5309	24	ABL33736
c	25	52	4.7	6636	24	ABN80023
c	26	52	4.7	7459	24	ABK31382
c	27	51.8	4.7	641	24	ABQ56694
c	28	51.6	4.7	47108	24	ABK31511
c	29	51.4	4.7	4041	21	AAA70170
c	30	51.2	4.7	3738	21	AAA70178
c	31	51.2	4.7	5690	22	AAS45368
c	32	51.2	4.7	5690	24	ABK28205
c	33	51.2	4.7	23695	24	ABQ66981
c	34	51	4.6	10286	22	AAS45309
c	35	51	4.6	10286	24	ABK28148
c	36	50.8	4.6	5880	22	AAS46331
c	37	50.8	4.6	5880	24	ABK28177
c	38	50.8	4.6	6254	24	ABL33621
c	39	50.6	4.6	12405	22	AAS45330
c	40	50.6	4.6	12405	24	AAS61143
c	41	50.6	4.6	12405	24	ABK28169
c	42	50.4	4.6	5689	22	AAS45384
c	43	50.4	4.6	5689	22	AAS46426
c	44	50.4	4.6	5689	24	ABK28226
c	45	50.4	4.6	8085	22	AAS46479

ALIGNMENTS

RESULT 1
AAA59599
ID AAA59599 standard; DNA; 6888 BP.
XX
AC
AC
XX
XX
DT 14-NOV-2000 (first entry)
XX
DE DNA encoding a cytochrome P450 enzyme designated DWF4.
DE
XX
KW DWF4; cytochrome P450 enzyme; brassinosteroid; 22alpha-hydroxylation;
KW plant phenotype; cell elongation; ss.
XX
OS Arabidopsis sp.
XX
FH Key
FH promoter
FT Location/Qualifiers
FT 1..3203
FT /tag= a
FT TATA_signal 3060..3125
FT /tag= b
FT CDS 3203..6110
FT /tag= c
FT /product= "DWF4"
FT /note= "contains introns"
FT exon 3203..3423
FT /tag= d
FT intron 3424..3503
FT /tag= e
FT exon 3504..3828
FT /tag= f
FT intron 3829..3913
FT /tag= g

FT exon 3914..4066
FT intron /*tag= h
FT exon 4067..4164
FT intron /*tag= i
FT exon 4165..4479
FT intron /*tag= j
FT exon 4480..4631
FT intron /*tag= k
FT exon 4632..4724
FT intron /*tag= l
FT exon 4725..4815
FT intron /*tag= m
FT exon 4816..4894
FT intron /*tag= n
FT exon 4895..5000
FT intron /*tag= o
FT exon 5001..5110
FT intron /*tag= p
FT exon 5111..5864
FT intron /*tag= q
FT exon 5865..6110
FT intron /*tag= r
FT exon 6011..6468
FT intron /*tag= s

W0200047715-A2.

XX 17-AUG-2000.
XX 11-FEB-2000; 2000WO-US03820.
XX 11-FEB-1999; 99US-0119657.
XX 11-FEB-1999; 99US-0119658.

(ARIZ-) ARIZONA BOARD OF REGENTS.

PI Aspiroz R, Choe S, Feldmann KA;

XX WPI; 2000-549142/50.
XX P-PSDB; AAB07921.

PT New isolated dwf4 polynucleotide useful for altering the phenotype of
PT plants, for diagnostic assays and in the production of antibodies -

XX Claim 3; Fig 10A-G; 113pp; English.

CC The present sequence encodes a DWF4 polypeptide. The polypeptide is a
CC cytochrome P450 enzyme that mediates multiple steps in synthesis of
CC brassinosteroids. Specifically, it mediates multiple
CC 2alpha-hydroxylation steps in brassinosteroid biosynthesis. The DWF4
CC polynucleotide is used for altering the phenotype of a plant. DWF4
CC plants display a dramatic reduction in the length of different organs,
CC and this size reduction is attributable to a defect in cell elongation.
CC The DWF4 polynucleotides and polypeptides can be used in diagnostic
CC assays and to generate antibodies, which can be used to produce
CC immunogenic compositions.

XX Sequence 6888 BP; 2294 A; 1010 C; 1193 G; 2391 T; 0 other;

Query Match 100.0%; Score 1101; DB 21; Length 6888;
Best Local Similarity 100.0%; Pred. No. 1.6e-208;
Matches 1101; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 AATCTACAAATTTATATTTAGTCAATACCAATGCATAGAAAGTTCCAAAAAATTT 60
DB 2102 AATCTACAAATTTATATTTAGTCAATACCAATGCATAGAAAGTTCCAAAAAATTT 2161

QY 61 TGTTACAGAACTTCCAAATTTTTTTTTTTATGGAACAAGAAATACAGATAGAAAC 120
DB 2162 TGTTACAGAACTTCCAAATTTTTTTTTTTATGGAACAAGAAATACAGATAGAAAC 2221

QY 121 TATTTTGTCTGGAATGGAAGTAGTATATATACATTAGCAAAATTTAAAAAATATATAA 180
TT

DB 2222 TATTTTGTCTGGAATGGAAGTAGTATATATACATTAAAGCAAAATTTAAAAAATATATAA 2281
QY 181 GCCTATACGCGCTCAAAAGTATCTTATCTAGTAGGTGTAATTAATATGATGCTGGGATT 240
DB 2282 GCCTATACGCGCTCAAAAGTATCTTATCTAGTAGGTGTAATTAATATGATGCTGGGATT 2341
QY 241 CAGAAATGGGACAAACAATGAAACGGAATTAATAATTAATTAACCTTTAAAAATAAATAAAT 300
DB 2342 CAGAAATGGGACAAACAATGAAACGGAATTAATAATTAATTAACCTTTAAAAATAAATAAAT 2401
QY 301 TGAGTAAATGTCTTTCTGACTATTGAGGGGCAAAAAAGACAATGCCAAAGCTCTAG 360
DB 2402 TGAGTAAATGTCTTTCTGACTATTGAGGGGCAAAAAAGACAATGCCAAAGCTCTAG 2461
QY 361 GGTCTGACTGTCCAGTTCGGTAATAATCTTAATACTCTCTTTGACCCACGCTCGTGT 420
DB 2462 GGTCTGACTGTCCAGTTCGGTAATAATCTTAATACTCTCTTTGACCCACGCTCGTGT 2521
QY 421 AGGGTCTCTTGACATTTTCTACTCTTACCCCTACTCTGTCGAGCCCACTTTTCCCAT 480
DB 2522 AGGGTCTCTTGACATTTTCTACTCTTACCCCTACTCTGTCGAGCCCACTTTTCCCAT 2581
QY 481 ATCCTAAGGGTAATTTTGGAAATCCCAATTTAAACCGGATTGAGACCGTACCGACTTCT 540
DB 2582 ATCCTAAGGGTAATTTTGGAAATCCCAATTTAAACCGGATTGAGACCGTACCGACTTCT 2641
QY 541 GGGATTCTGTCGAGCATTTTATCAAAAATTTATTAGCACGAATGGGTTTATTAATTTAAAA 600
DB 2642 GGGATTCTGTCGAGCATTTTATCAAAAATTTATTAGCACGAATGGGTTTATTAATTTAAAA 2701
QY 601 ACTCACAACCTTGATCAGATAAAATTTTCAATAACACTTTTACGATGGATTTCGACGATCA 660
DB 2702 ACTCACAACCTTGATCAGATAAAATTTTCAATAACACTTTTACGATGGATTTCGACGATCA 2761
QY 661 TCTAATGACTTTTTTTTTTCTACCACGGTGGATGAAAGTTATAGTACTATTAGCCAGAGA 720
DB 2762 TCTAATGACTTTTTTTTTTCTACCACGGTGGATGAAAGTTATAGTACTATTAGCCAGAGA 2821
QY 721 CAATTGATTATAGATATATCCATTAATCCATGATATTTATGATATATAATAGCTGTTAAAC 780
DB 2822 CAATTGATTATAGATATATCCATTAATCCATGATATTTATGATATATAATAGCTGTTAAAC 2881
QY 781 TATTTACGATCGCAGCTTTCTGCAACTTTTCTTTTATTTAATTTAAGAGTTTAAATAAATAA 840
DB 2882 TATTTACGATCGCAGCTTTCTGCAACTTTTCTTTTATTTAAGAGTTTAAATAAATAA 2941
QY 841 AGTATTAAGGAGCATTAACGAGGCAACAAAAGTAAATGAACACGAGAAACAAAGCCAT 900
DB 2942 AGTATTAAGGAGCATTAACGAGGCAACAAAAGTAAATGAACACGAGAAACAAAGCCAT 3001
QY 901 GAAGCTCATTTGTTAGTTTAAAGCTTAATAAGAGATTTTATTAATTTAATGACGATGA 960
DB 3002 GAAGCTCATTTGTTAGTTTAAAGCTTAATAAGAGATTTTATTAATTTAATGACGATGA 3061
QY 961 TAACAATTTATTTTCTGACTTTTAAACCCCTCTTTACAAACAGAGCTCCCTTTT 1020
DB 3062 TAACAATTTATTTTCTGACTTTTAAACCCCTCTTTACAAACAGAGCTCCCTTTT 3121
QY 1021 CAGTAGAAGTCCGATTTCCCAATCTTAAGACAAACCCATTAGAAAGAGAAAGTGGAG 1080
DB 3122 CAGTAGAAGTCCGATTTCCCAATCTTAAGACAAACCCATTAGAAAGAGAAAGTGGAG 3181
QY 1081 AGAGAGAGAGAAACTAGCTCC 1101
DB 3182 AGAGAGAGAGAACTAGCTCC 3202

RESULT 2

ABL70390/c

ID ABL70390 standard; DNA; 6107 BP.

XX ABL70390;

XX ABL70390;

DT 01-JUL-2002 (first entry)
XX Chemically treated cell signalling DNA sequence complementary to#140.
XX Cell signalling; cytosine methylation; cell signalling disease;
KW cancer; tumour; cytostatic; ds.
XX Unidentified.
XX WO200202807-A2.
XX 10-JAN-2002.
XX 29-JUN-2001; 2001WO-EP07471.
XX 30-JUN-2000; 2000DE-1032529.
PR 01-SEP-2000; 2000DE-1043826.
XX (EPIG-) EPIGENOMICS AG.
PA Olek A, Piepenbrock C, Berlin K;
PI WPI; 2002-154758/20.
DR Nucleic acid, useful for diagnosis and therapy of diseases associated
XX with cell signalling e.g. cancer, comprises chemically modified genomic
XX sequences of genes associated with cell signalling.
XX Claim 1; SEQ ID NO 280; 24ppt+sequence listing; English.
XX The invention relates to a nucleic acid comprising a sequence of at least
CC 18 bases of a segment of chemically pretreated DNA of genes associated
CC with cell signalling. The activity of the modified sequences of the
CC invention may be described as cytostatic. The object of the invention is
CC to provide the chemically modified DNA of genes associated with cell
CC signalling, as well as oligonucleotides and/or pNA-oligomers for
CC detecting cytosine methylations, as well as a method which is
CC particularly suitable for the diagnosis and/or therapy of genetic and
CC epigenetic parameters of genes associated with cell signalling. The
CC chemically modified DNA provided by the invention is useful for diagnosis
CC and therapy of diseases such as solid tumours and cancer. The sequences
CC given in records ABL70111-ABL70626 represent chemically pre-treated
CC genomic DNA's of genes associated with cell signalling.
CC Note: The sequence data for this patent is not represented in the printed
CC specification, but is based on sequence information supplied by the
CC European Patent Office.
XX
SQ Sequence 6107 BP; 1604 A; 136 C; 1251 G; 3116 T; 0 other;
Query Match 5.1%; Score 56.6; DB 24; Length 6107;
Best Local Similarity 47.8%; Pred. No. 0.044;
Matches 164; Conservative 0; Mismatches 179; Indels 0; Gaps 0;
QY 26 CAAATACATGATGACAAAGTTCCCAAAAAATTTTGGTAAAGAAACCTCCAAATTTT 85
Db 4771 CAAAAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATCCT 4712
QY 86 TTTTTTATGGACAGAAATACAGATAGAAACATTTTGTGTGGATGGAAGTACT 145
Db 4711 TTAATATAAACTAAAAAACCCTAAATTTTATATATACAAAAATCAAAATTTT 4652
QY 146 AATATACATTAAGCAATTTTAAAAAATTTATATAGCCCTACGCGCTCAAAAGTATGTTA 205
Db 4651 AAACCTTAACTATATAAATCATATATACTAAAAAAGCTAAAAAATAAATAAATAA 4592
QY 206 TCTAGTAGTGTAATTAATGATGCGGATTCAGAAATGGGACACAAATGAAACG 265
Db 4591 TAAATATAAATAAATAAATAAATCATATATATATATATATATATATATATATATAT 4532
QY 266 GAATTAATAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 325
Db 4531 TAAACTATAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 4472

QY 326 GAGGGGCAAAAAAGACAAATGCCAAAGTCTACGGTTTGAC 368
Db 4471 TTTATCAAAATAAATCATTTTCAAAATAATACATTTATAC 4429
RESULT 3
AAS61342/C
ID AAS61342 standard; DNA; 6107 BP.
XX AAS61342;
AC AAS61342;
XX 29-JAN-2002 (first entry)
XX Human gene regulation-associated gene oligonucleotide #297.
XX Human; Gene regulation-associated gene; severe combined immunodeficiency;
KW cardiac damage; inflammatory response; Haemophilia; Werner syndrome;
KW asthma; HDR syndrome; congenital heart defect; Saethre-Chotzen syndrome;
KW renal disease; Preeclampsia; cardiac allograft vascular disease;
KW colorectal cancer; thyroid cancer; oesophageal cancer; ds; tumour;
KW immunostimulant; cardiant; antiinflammatory; coagulant; antiasthmatic;
KW nephrotropic; gynecological; anti-tumour; immunosuppressive; cytostatic.
XX
OS Homo sapiens.
XX WO200177375-A2.
XX 18-OCT-2001.
XX 06-APR-2001; 2001WO-EP03968.
XX 06-APR-2000; 2000DE-1019058.
PR 07-APR-2000; 2000DE-1019173.
PR 30-JUN-2000; 2000DE-1032529.
PR 01-SEP-2000; 2000DE-1043826.
XX (EPIG-) EPIGENOMICS AG.
PA Olek A, Piepenbrock C, Berlin K;
PI WPI; 2002-017470/02.
XX New nucleic acid sequences from chemically modified genes associated
PT with gene regulation, useful for analysing cytosine methylations for
PT diagnosis and therapy of diseases e.g. severe combined immunodeficiency
PT disease
XX
PS Disclosure; SEQ ID NO 304; 26pp; English.
XX The invention relates to 224 nucleic acid sequences comprising at least
CC 18 bases of a chemically pretreated gene associated with gene regulation
CC selected from 43 known genes (or complementary sequences). The
CC chemical pretreatment converts cytosine bases unmethylated at the
CC 5-position to uracil or another base with hybridisation behaviour
CC dissimilar to cytosine, to enable analysis of cytosine methylations.
CC The DNA sequences, oligomers (or sets/arrays) and method are
CC useful in the diagnosis of diseases (or predisposition to diseases)
CC associated with gene regulation and in therapy of such diseases; by
CC enabling analysis of the cytosine methylation patterns of such genes,
CC kits are provided. They are especially useful in diagnosis
CC and therapy of e.g. severe combined immunodeficiency disease, cardiac
CC disorders, haemophilia, solid tumours and cancer, Werner syndrome,
CC asthma, HDR syndrome, Saethre-Chotzen syndrome, renal disease,
CC preeclampsia, graft versus-host disease. The present sequence is a
CC sequence included in the sequence data for this specification and is
CC associated with the human gene regulation-associated genes.
CC Note: The sequence data for this patent did not form part
CC of the printed specification, but was obtained in electronic
CC format directly from WIPO at
CC ftp.wipo.int/pub/published_pct_sequences
XX
SQ Sequence 6107 BP; 1604 A; 136 C; 1251 G; 3116 T; 0 other;

Query Match	5.1%	Score 56.6	DB 24	Length 6107
Best Local Similarity	47.8%	Pred. No. 0.044		
Matches 164	Conservative 0	Mismatches 179	Indels 0	Gaps 0
QY	26	CAATAACAATGCATGACAAAGTTCCRAAAAAAAAAATTTTGTTAACAGAAACCTTCACAAATTTT	85	
DB	4771	CAAAAAAAAAATAAATAAATAATATAATAATAAAAAATCTAAAAATACATAAAAAATATCCCT	4712	
QY	86	TTTTTTTTATGGACACAGAAATACACAGATAGAAAACCTATTTTTGTGTGGAAATGGAAGTACT	145	
DB	4711	TTAAATATAAAAACTAAAAAANAACCACCTAAAAATTTATAATTTACAAAAATCAAAAAATTTT	4652	
QY	146	AATATACATTAAGCAAAATTTTAAAAAATPATATAAGCCTATACGGGCTCAAAAGTAGTTA	205	
DB	4651	AAACTTAACATATATAAAACCTCATATAACCTAAAAAACAAGTAAAAAATAAACTAAAAATAAAA	4592	
QY	206	TCTAGTAGGTGAATTAATAATGCATGGTGGCATTACAGAAATTTGGGCACACAATGAAAACG	265	
DB	4591	TAAATATAAAAAAATAAATTAATTCATATATATCATATATCAACAAAAATTAATAATCTATCAAAAT	4532	
QY	266	GAATTAAAAATTAACCTTTAAAAATAAATAAAAAATTTGAGTAATAATGTGTTTCTCGACTATT	325	
DB	4531	TAAACTCATATTTTAAAAAANAATACGAAAACTACTATTATTATATATCACTTTTCTTTTCC	4472	
QY	326	GAGGGGCAAAAAAGACAAATGCCAAAAGCTACGGGTTTGAC	368	
DB	4471	TTTAAATCAAAATAAAATCAATTTCAAAAAATAATACATTTTATAAC	4429	

RESULT 4	
ABK31431/c	
ID	ABK31431 standard; DNA; 6107 BP.
XX	
XX	
AC	ABK31431;
XX	
XX	
DT	23-APR-2002 (first entry)
XX	
XX	Signal transduction associated gene modified complementary DNA #137.
DE	
XX	
KW	Human; signal transduction associated gene; cytosine methylation state;
KW	CpG island; signal transduction associated disease; solid tumour; cancer;
KW	antitumour; cytostatic; mutant; ds.

XXA Oligonucleotide for diagnosis and therapy of diseases associated with
XXB signal transduction e.g. cancer, comprises chemically modified genomic
XXC sequences of genes associated with signal transduction -
XXD
XXE Claim 1; SEQ ID No 274; 24pp; English.
PS

	epigenetic parameters of genes associated with signal transduction.
	The genomic DNA can be obtained from cells or cellular components which
	contain DNA, e.g. cell lines, biopsies, blood, sputum, stool, urine,
	cerebral-spinal fluid, tissue embedded in paraffin such as tissue from
	eyes, intestine, kidney, brain, heart, prostate, lung, breast or liver,
	histologic object slides, and all their possible combinations. The
	sequences of the invention are useful for the diagnosis and therapy of
	diseases associated with signal transduction e.g. solid tumours and
	cancer. ABK31158-ABK31545 represent chemically pretreated genomic DNA
	sequences of different genes associated with signal transduction, or
	their complementary sequences.
	Note: The sequence data for this patent did not form part of the printed
	specification, but was obtained in electronic format directly from the
	European Patent Office.
XX	
SQ	Sequence 6107 BP; 1604 A; 136 C; 1251 G; 3116 T; 0 other;
	Query Match 5.1%; Score 56.6; DB 24; Length 6107;
	Best Local Similarity 47.8%; Pred. No. 0.044;
	Matches 164; Conservative 0; Mismatches 179; Indels 0; Gaps 0;
QY	26 CAATACAACTGATAGAAAGTTCCAAAAAATAATTTTGTTTAAACAGAAAACCTTCAAAATTTTT 85
Db	4771 CAAAAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAA 4712
QY	86 TTTTNTTTATGGCAACAAGAATAAACAGATAGAAAACACTATTCTTGTGGAATGGAAGTAGT 145
Db	4711 TTAAATAAAAAACTAAAAAANAACCACATAAAAAATTTTAATTTACAAAAATCAAAAATTTT 4652
QY	146 AATATACATTAAGCAAATTTTAAAAAATTTATATAAGCCTATACGCCGCTCAAGATGTGTA 205
Db	4651 AAACCTTAACATATAAAACTCATATAAAGTAAAAAACAAGTAAAAAATAAAGTAAAAAATA 4592
QY	206 TCTAGTAGGTGTAANTTAATAATGCATGGTGGCGATTTCAGAAATTGGGACCAACAAATGAAACG 265
Db	4591 TAAAAATAAAAAAATAAATAAATTTTCATATATCATATAAACAATAAATAAATAAATAAATA 4532
QY	266 GAATTTAAATAATATAACTTTTAAAAATAAATAAATAAATAAATAAATAAATAAATAAATA 325
Db	4531 TAAAACTCATAAATTTTAAAAAANAATAACGNAACACTATTATATATATCACTTTTCTTTCC 4472
QY	326 GAGGGGCAAAAAAGACAAATGCCAAAAAGTCTACGGGTTTGAC 368
Db	4471 TTTTAATCAAAATAAATAATCATTTCCAATAATAATACATATTTATAC 4429

RESULT 5	
AAA01920	
ID	AAA01920 standard; cDNA; 875 BP.
XX	
XX	AAA01920;
XX	
DT	19-MAY-2000 (first entry)
XX	
DE	Human colon cancer cell line polynucleotide sequence SEQ ID NO:1911.
XX	
XX	Human; colon cancer; tumour; diagnosis; gene expression product;
KW	probe; detection; cancerous stage; metastasis; identification;
KW	breast cancer; oestrogen receptor-positive breast cancer; therapy;
KW	oestrogen receptor-negative breast cancer; lung cancer; ss.
XX	
XX	Homo sapiens.
XX	
PN	W09958675-A2.
XX	
XX	18-NOV-1999.
XX	
XX	13-MAY-1999; 99WO-US10602.
XX	
PR	14-MAY-1998; 98US-0085426.
PR	15-MAY-1998; 98US-0085537.
PR	15-MAY-1998; 98US-0085696.
PR	21-OCT-1998; 98US-0105234.

```

PR 27-OCT-1998; 98US-0105877.
XX (CHIR) CHIRON CORP.
PA (HYSE-) HYSEQ INC.
XX
PI Williams LT, Escobedo J, Innis MA, Garcia PD, Sudduth-Klinger J;
PI Reinhard C, Giese K, Randazzo F, Kennedy GC, Pot D, Kassam A;
PI Lamson G, Drmanac R, Crkvenjakov R, Dickson M, Drmanac S, Labat I;
PI Leshkowitz D, Kita D, Garcia V, Jones LW, Stache-Crain B;
XX WPI; 2000-126369/11.
XX
XX Polynucleotide library used to determine cancerous states of mammalian
XX cells -
XX
XX Claim 1; Page 720; 1097pp; English.
XX
XX AAA00010 to AAA02716 represent polynucleotides isolated from cDNA
XX libraries constructed from human colon cancer cell lines. The present
XX invention also describes a method of detecting differentially expressed
XX genes correlated with a cancerous state of a mammalian cell, comprising
XX detecting at least one differentially expressed gene product in a test
XX sample derived from a cell suspected of being cancerous, where detection
XX of the differentially expressed gene product is correlated with a
XX cancerous state of the cell from which the test sample was derived.
XX The polynucleotides sequences can be used in a method for detecting
XX differentially expressed genes correlated with a cancerous state of a
XX mammalian cell. The polynucleotides can also be used as probes for a
XX detecting and mapping related genes. They can be used in diagnosis and
XX prognosis of diseases and disorders (e.g. identification of
XX pre-metastatic or metastatic cancerous states, stages of cancer, or
XX responsiveness of cancer to therapy). This is particularly for breast
XX cancer, oestrogen receptor-positive breast cancer, oestrogen receptor-
XX negative breast cancer, lung cancer, and colon cancer.
XX
XX Sequence 875 BP; 421 A; 10 C; 78 G; 198 T; 168 other;
XX
XX Query Match 4.9%; Score 54.2; DB 21; Length 875;
XX Best Local Similarity 42.6%; Pred. No. 0.1;
XX Matches 149; Conservative 0; Mismatches 196; Indels 5; Gaps 1;
XX
XX 1 AATCTACAAATTTATTAATATTTAGTCAATCAATGAGTGAAGTTCACAAAATTT 60
XX 525 AATANTNATATATANTANNAATAATATAATAAATAAGAGATATGNNAAANGAAAAA 584
XX
XX 61 TGTACAGAACTCCAAATTTTATTTTATGGAACAAGAAATAACAGATAGAAAC 120
XX 585 ATATANAGGAAAGTAAATTAATNGATATTTAAGANAAAGAAANAAATATANAN 644
XX 121 TATTTTGTGTCGAATGGAAGTAGTAAATATACATTAAGCAATTTTAAATAATTAATAA 180
XX 645 NATNANAATATANTANNTANNAATAANANAAANATNCNAATNTANNAGATNATAAGA 704
XX 181 GCCTATACGGCTCA-----AAGTAGTGTATCTAGTAGGTGTAATTAATTAATGCGATGGTG 235
XX 705 NAANNTATNAATNAACNTGAAATATATNTAANNATAGNACTTANANTNTTATAAGAN 764
XX 236 CGATTCAGAAATGGGACACATGAAACGGGAATTAATAATTAATTAATAATAATAA 295
XX 765 GNNTTAAGANATATANATNAATAGATANNNTNAAATAAATTAATATATAAATAATANNA 824
XX 296 AAATTGTAGTAATGTGTTTCTGACTATTGAGGGGCAAAAGACAA 345
XX 825 AATAATGAGTGANAGATATANNANNTANNTANNGTATNTAATAATAA 874
XX
XX RESULT 6
XX ABL32803/c
XX ID ABL32803 standard; DNA; 6113 BP.
XX AC ABL32803;
XX XX
XX DT 26-MAR-2002 (first entry)

```

```

XX Human immune system associated gene SEQ ID NO: 776.
XX DE
XX XX
XX KW Human; immune system disease; cytosine methylation; antiasthmatic;
XX antiarteriosclerotic; antianaemic; cytostatic; nootropic;
XX neuroprotective; anti-HIV; anticonvulsant; ophthalmological;
XX antirheumatic; antiarthritis; antidiabetic; antipsoriatic;
XX antileukematory; cancer; eye disease; arteriosclerosis; anaemia;
XX acute myeloid leukaemia; Alzheimer's disease; AIDS; epilepsy;
XX neurofibromatosis; rheumatoid arthritis; psoriasis; bowel disease;
XX gene; ds.
XX
XX Homo sapiens.
XX OS
XX XX
XX PN WO200200928-A2.
XX XX
XX PD 03-JAN-2002.
XX XX
XX PP 02-JUL-2001; 2001WO-EP07537.
XX XX
XX PR 30-JUN-2000; 2000DE-1032529.
XX XX
XX PR 01-SEP-2000; 2000DE-1043826.
XX XX
XX PA (EPIG-) EPIGENOMICS AG.
XX XX
XX PI Olek A, Piepenbrock C, Berlin K;
XX XX
XX DR WPI; 2002-130909/17.
XX XX
XX PT Nucleic acid comprising fragment of chemically modified gene, useful
XX for diagnosis and treatment of diseases associated with abnormal
XX cytosine methylation
XX
XX PS Claim 1; SEQ ID NO 776; 32pp + Sequence Listing; German.
XX XX
XX CC The present invention provides a number of human immune system associated
XX genes which are modified by the methylation of cytosines. The sequences
XX can be used in the diagnosis and treatment of immune system disorders,
XX including eye diseases such as retinopathy, neovascular glaucoma and
XX macular degeneration, arteriosclerosis, anaemia, cancer, acute myeloid
XX leukaemia, Alzheimer's disease, AIDS, epilepsy, neurofibromatosis,
XX rheumatoid arthritis, psoriasis and inflammatory/ulcerative bowel
XX diseases. The present sequence is a gene of the invention.
XX
XX SQ Sequence 6113 BP; 1694 A; 157 C; 1445 G; 2817 T; 0 other;
XX
XX Query Match 4.9%; Score 54.2; DB 24; Length 6113;
XX Best Local Similarity 51.9%; Pred. No. 0.13;
XX Matches 122; Conservative 0; Mismatches 113; Indels 0; Gaps 0;
XX
XX 6 ACAATATTATTAATTTAGTCAATAACAATGCATGAAGTTCACAAAATTTTGTGA 65
XX 5823 ACACAAACTAAATAAATAAATAAACAATACTTAAATACTCAATAAATAATATCA 5764
XX
XX 66 ACAGAACTCCAAATTTTATTTTATGGAACAAGAAATAACAGATAGAAATATTT 125
XX 5763 AAAAAAATTTAAATATATATATATAAATAAATAAATAATTTATATATATTTAT 5704
XX 126 TGTGTGGAATGGAAGTAGTAATACATTAACCAATTTTAAATAATATAAGCCTA 185
XX 5703 TCCTTTTAAATAAATTCATAAATAAATAAATAAATAATTTTATTTATACGTATCA 5644
XX 186 TAGCGCTCAAGTAGTGTATCTAGTAGGTGTAATTAATTAATGATGCGGATT 240
XX 5643 CGTATCTATAAATCTTTTTTTTTTTTTTTTTTTTTTAAATAAATCTCGCCCTATT 5589
XX
XX RESULT 7
XX ABL34197/c
XX ID ABL34197 standard; DNA; 37973 BP.
XX AC ABL34197;
XX XX

```


[illegible]

Query Match 4.8%; Score 52.6; DB 24; Length 16228;
Best Local Similarity 48.2%; Pred. No. 0.31;
Matches 148; Conservative 0; Mismatches 159; Indels 0; Gaps 0;

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: June 10, 2003, 22:02:58 ; Search time 52.4398 Seconds
(without alignments)
6438.831 Million cell updates/sec

Title: US-09-502-426B-1_COPY_2102_3202

Perfect score: 1101

Sequence: 1 aatcacaaattattaat.....gagagagagaactagctcc 1101

Scoring table: IDENTITY_NUC

Gapop 10.0 , Gapext 1.0

Searched: 441362 seqs, 15338381 residues

Total number of hits satisfying chosen parameters: 882724

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

Issued_Patents_NA:*
1: /cgn2.6/ptodata/1/lna/5A-COMB.seq:*
2: /cgn2.6/ptodata/1/lna/5B-COMB.seq:*
3: /cgn2.6/ptodata/1/lna/6A-COMB.seq:*
4: /cgn2.6/ptodata/1/lna/6B-COMB.seq:*
5: /cgn2.6/ptodata/1/lna/PCTUS-COMB.seq:*
6: /cgn2.6/ptodata/1/lna/backfiles1.seq:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
C 1	49.2	4.5	3350	3	US-08-617-860B-3
C 2	49.2	4.5	6265	4	US-09-129-112-3
C 3	46.6	4.2	19124	2	US-08-487-826B-13
C 4	46.4	4.2	636	4	US-08-998-416-1137
C 5	46	4.2	731	1	US-08-451-405A-2
C 6	45.8	4.2	665	2	US-08-883-795A-36
C 7	45.8	4.2	51952	3	US-08-947-823-1
C 8	45	4.1	20674	4	US-09-641-638-651
C 9	44.8	4.1	3138	1	US-07-867-106-4
C 10	44.6	4.1	19124	2	US-08-487-826B-13
C 11	44	4.0	7218	1	US-08-232-463-14
C 12	43.8	4.0	6124	4	US-08-213-419B-3
C 13	43.6	4.0	5852	1	US-07-867-106-2
C 14	43.2	3.9	660	1	US-07-991-867B-32
C 15	43.2	3.9	660	1	US-08-107-755A-32
C 16	43.2	3.9	660	2	US-08-544-332-32
C 17	43.2	3.9	660	4	US-08-370-861A-32
C 18	43.2	3.9	1511	1	US-07-991-867B-8
C 19	43.2	3.9	1511	1	US-08-107-755A-8
C 20	43.2	3.9	1511	1	US-08-544-332-8
C 21	43.2	3.9	1511	4	US-09-370-861A-8
C 22	43.2	3.9	4810	3	US-08-852-629-11
C 23	43.2	3.9	4838	3	US-08-852-629-15
C 24	42.8	3.9	1850	3	US-08-617-860B-32
C 25	42.8	3.9	4098	2	US-08-605-106-4
C 26	42.4	3.9	470	4	US-09-020-956-102
C 27	42.4	3.9	470	4	US-09-030-607-102

C 28	42.4	3.9	470	4	US-09-605-785-102	Sequence 102, App
C 29	42.4	3.9	470	4	US-09-439-313-102	Sequence 102, App
C 30	42.4	3.9	470	4	US-09-352-616A-102	Sequence 102, App
C 31	42.4	3.9	470	4	US-09-232-149A-102	Sequence 102, App
C 32	42.4	3.9	615	4	US-08-998-416-186	Sequence 186, App
C 33	42.4	3.9	5181	1	US-08-257-073-10	Sequence 10, Appli
C 34	42.4	3.8	1406	3	US-08-913-842-6	Sequence 6, Appli
C 35	41.8	3.8	658	4	US-08-998-416-595	Sequence 595, App
C 36	41.4	3.8	2861	1	US-08-299-953-1	Sequence 1, Appli
C 37	41.4	3.8	2861	1	US-08-459-415-1	Sequence 1, Appli
C 38	41.4	3.8	2861	4	US-09-066-687-1	Sequence 1, Appli
C 39	41.4	3.8	2861	5	PCT-US95-11231-1	Sequence 1, Appli
C 40	41.4	3.8	3881	1	US-08-299-953-2	Sequence 2, Appli
C 41	41.4	3.8	3881	1	US-08-459-415-2	Sequence 2, Appli
C 42	41.4	3.8	3881	4	US-09-066-687-2	Sequence 2, Appli
C 43	41.4	3.8	3881	5	PCT-US95-11231-2	Sequence 63, Appli
C 44	41	3.7	1441	4	US-08-821-994-63	Sequence 1, Appli
C 45	41	3.7	2110	4	US-09-419-459-1	Sequence 1, Appli

ALIGNMENTS

RESULT 1
US-08-617-860B-3/c
; Sequence 3, Application US/08617860B
; Patent No. 6133506
; GENERAL INFORMATION:
; APPLICANT: Tvpfer, R., Bautor, J., Bothmann, H., Filsak, E.,
; APPLICANT: Hvricke-Grandpierre, C., Klein, B., Martini, N.,
; APPLICANT: M ller, A., Schulte, W., Voetz, M., Walek, J.,
; APPLICANT: Schell, J.
; TITLE OF INVENTION: Promoters
; NUMBER OF SEQUENCES: 35
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Steinberg, Raskin & Davidson, P.C.
; STREET: 1140 Avenue of the Americas
; CITY: New York
; STATE: New York
; COUNTRY: New York
; ZIP: 10036
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette, 3.50 inch, 1.4 Mb storage
; COMPUTER: IBM Compatible
; OPERATING SYSTEM: DOS
; SOFTWARE: Patent Release #1.0, Version #1.25 (EPO)
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/617,860B
; FILING DATE: 01-MAR-1996
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: PCT/EP94/02950
; FILING DATE: 05-SEP-1994
; APPLICATION NUMBER: DE P4329951.2
; FILING DATE: 04-SEP-1993
; INFORMATION FOR SEQ ID NO: 3:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 3350 Base pairs
; TYPE: Nucleic acid
; STRANDEDNESS: Double stranded
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
; HYPOTHETICAL: NO
; ANTI-SENSE: NO
; ORIGINAL SOURCE:
; ORGANISM: Brassica napus
; IMMEDIATE SOURCE:
; LIBRARY: genomic Lambda FIX II
; CLONE: BnACCasg10
; FEATURE:
; NAME/KEY: Startcodon
; LOCATION: 2611..2613
; FEATURE:
; NAME/KEY: CDS

US-08-617-860B-3
; LOCATION: join(2611..2908, 3001..3341)

Query Match	4.5%;	Score	49.2;	DB	3;	Length	3350;
Best Local Similarity	48.7%;	Pred.	No.	0.02;			
Matches	190;	Conservative	0;	Mismatches	198;	Indels	2; Gaps
QY	6	ACAAATTAATTAATATTTAGCTCAATACAAATGCATACAGAGTCCCAAAAAAATTTTGTTA	65				
Db	2312	ATAAATGAATTAATTTATTAATTAATTAATATATAT-CAGCTATGTAATAAAAATATAGAA	2254				
QY	66	ACAGAAACTCCAAATTTTTTTTTTATGGACACAGAAATACAGATAGAAACATATTT	125				
Db	2253	ACTAAAAATTAATTAATTTAAAAATGTTTTAAAAACATAAATAATAGTTGGGTA-TGTAT	2195				
QY	126	TGTTGGGAATGGAAGTAGTAATATACATTAAGCAAAATTTAAAAAATTAATATAGCCCTA	185				
Db	2194	ATATTTTATGTTTAAAGACATTTAGAAAAATATTAATTTTAGTTTTTAAATTTTATTTTC	2135				
QY	186	TACGGCTCAAAGTAGTTATCTAGTAGGTGAATTAATAATGCATGGTCCGATTCAGAA	245				
Db	2134	ATTTGACATACAAAATATCAAAAAATAGTTAGACTATTATAAAAATTTCTAACAAAGT	2075				
QY	246	TTGGGACAAACATGAAACGGGAATTAANAATTAATTAACCTTTAAATAAATAAAAAATTTGAGT	305				
Db	2074	AAGAGTTATGATATATAAAAAAATCAAGACATAAAAATTTATAAATATTTAAAAATTTAAT	2015				
QY	306	AAATGTGTTTCTGACTATTGAGGGCAAAAAACACAAATGCCAAAGTCTACGGGTTT	365				
Db	2014	GTGAATAATAAATTAACCTTTAAATCAAAATAAACCAAAAAATAAAGATCATTTGAATA	1955				
QY	366	GACTGTCCAGTTCCGGTAATACTAATAAC	395				
Db	1954	AATTTGCAATAACAGAAATAAACTACACC	1925				

RESULT 2
 US-09-129-112-3/c
 ; Sequence 3, Application US/09129112
 ; Patent No. 6465716
 ; GENERAL INFORMATION:
 ; APPLICANT: Etzler, Marilynn B.
 ; APPLICANT: Murphy, Judith B.
 ; APPLICANT: The Regents of the University of California
 ; TITLE OF INVENTION: A No. 6465716 Factor Binding Protein From Legume Roots
 ; FILE REFERENCE: 023070-079810US
 ; CURRENT APPLICATION NUMBER: US/09/129,112
 ; CURRENT FILING DATE: 1998-08-04
 ; PRIOR APPLICATION NUMBER: US 08/907,226
 ; PRIOR FILING DATE: 1997-08-06
 ; NUMBER OF SEQ ID NOS: 19
 ; SOFTWARE: PatentIn Ver. 2.1
 ; SEQ ID NO 3
 ; LENGTH: 6265
 ; TYPE: DNA
 ; ORGANISM: Dolichos biflorus
 ; FEATURE:
 ; OTHER INFORMATION: genomic sequence of NEP46 (DB46)
 ; NAME/KEY: exon
 ; LOCATION: {633}..(944)
 ; NAME/KEY: intron
 ; LOCATION: {945}..(1022)
 ; NAME/KEY: exon
 ; LOCATION: {1023}..(1151)
 ; NAME/KEY: intron
 ; LOCATION: {1152}..(1559)
 ; NAME/KEY: exon
 ; LOCATION: {1560}..(1616)
 ; NAME/KEY: intron
 ; LOCATION: {1617}..(1697)
 ; NAME/KEY: exon
 ; LOCATION: {1698}..(1790)
 US-09-129-112-3

RESULT 3
 US-08-487-826B-13
 ; Sequence 13, Application US/08487826B
 ; Patent No. 5993827
 ; GENERAL INFORMATION:
 ; APPLICANT: Sim, Kim L.
 ; APPLICANT: Chitnis, Chetan
 ; APPLICANT: Miller, Louis H.
 ; APPLICANT: Peterson, David S.
 ; APPLICANT: Su, Xin-zhaun
 ; APPLICANT: Wellem, Thomas E.
 ; TITLE OF INVENTION: BINDING DOMAINS FROM PLASMODIUM VIVAX
 ; TITLE OF INVENTION: AND PLASMODIUM FALCIPARUM ERYTHROCYTE
 ; NUMBER OF SEQUENCES: 45
 ; CORRESPONDENCE ADDRESS:
 ; ADDRESSEE: Knobbe Martens Olson & Bear
 ; STREET: 620 Newport Center Drive 16th Floor
 ; CITY: Newport Beach
 ; STATE: California
 ; COUNTRY: US
 ; ZIP: 92660
 ; COMPUTER READABLE FORM:
 ; MEDIUM TYPE: Floppy disk
 ; COMPUTER: IBM PC compatible
 ; OPERATING SYSTEM: PC-DOS/MS-DOS
 ; SOFTWARE: PatentIn Release #1.0, Version #1.25
 ; CURRENT APPLICATION DATA:
 ; APPLICATION NUMBER: US/08/487.826B
 ; FILING DATE: 10-SEP-1993
 ; CLASSIFICATION: 435
 ; ATTORNEY/AGENT INFORMATION:
 ; NAME: Israelson, Ned
 ; REGISTRATION NUMBER: 29,655
 ; REFERENCE/DOCKET NUMBER: NH121.001CPI
 ; TELECOMMUNICATION INFORMATION:
 ; TELEPHONE: (619) 235-8550
 ; TELEFAX: (619) 235-0176
 ; INFORMATION FOR SEQ ID NO: 13:
 ; SEQUENCE CHARACTERISTICS:
 ; LENGTH: 19124 base pairs
 ; TYPE: nucleic acid
 ; STRANDEDNESS: single
 ; TOPOLOGY: linear

RESULT 5
 US-08-451-405A-2
 ; Sequence 2, Application US/08451405A
 ; Patent No. 5736358
 ; GENERAL INFORMATION:
 ; APPLICANT: FASEL, NICOLAS JOSEPH
 ; APPLICANT: REYMOND, CHRISTOPHE DOMINIQUE
 ; TITLE OF INVENTION: DICTYOSTELID EXPRESSION VECTOR AND
 ; TITLE OF INVENTION: METHOD FOR EXPRESSING A DESIRED PROTEIN
 ; NUMBER OF SEQUENCES: 3
 ; CORRESPONDENCE ADDRESS:
 ; ADDRESSEE: THE WEBB LAW FIRM
 ; STREET: 700 KOPPERS BUILDING, 436 SEVENTH AVENUE
 ; CITY: PITTSBURGH
 ; STATE: PENNSYLVANIA
 ; COUNTRY: UNITED STATES OF AMERICA
 ; ZIP: 15219-1818
 ; COMPUTER READABLE FORM:
 ; MEDIUM TYPE: 3.5" FLOPPY DISK
 ; COMPUTER: Midwest Micro 486-50
 ; OPERATING SYSTEM: DOS
 ; SOFTWARE: WORDPERFECT 6.1
 ; CURRENT APPLICATION DATA:
 ; APPLICATION NUMBER: US/08/451,405A
 ; FILING DATE: 26-MAY-1995
 ; CLASSIFICATION: 435
 ; PRIOR APPLICATION DATA:
 ; APPLICATION NUMBER: 07/965,273
 ; FILING DATE: 15-JAN-1993
 ; INFORMATION FOR SEQ ID NO: 2:
 ; SEQUENCE CHARACTERISTICS:
 ; LENGTH: 731
 ; TYPE: NUCLEIC ACID
 ; STRANDEDNESS: SINGLE
 ; TOPOLOGY: UNKNOWN
 ;-08-451-405A-2
 Query Match 4.2%; Score 46; DB 1; Length 731;
 Best Local Similarity 46.3%; Pred. No. 0.07;


```

RESULT 8
US-09-641-638-651/c
; Sequence 651, Application US/09641638
; Patent No. 6432648
; GENERAL INFORMATION:
; APPLICANT: Blumenfeld, Marta
; APPLICANT: Bougueleret, Lydie
; APPLICANT: Chumakov, Ilya
; APPLICANT: Cohen, Annick
; TITLE OF INVENTION: BIALLELIC MARKERS DERIVED FROM GENOMIC REGIONS CARRYING
; TITLE OF INVENTION: GENES INVOLVED IN ARACHIDONIC ACID METABOLISM
; FILE REFERENCE: GENSET.051CP1
; CURRENT APPLICATION NUMBER: US/09/641,638
; CURRENT FILING DATE: 2000-08-16
; PRIOR APPLICATION NUMBER: US/09/502,330
; PRIOR FILING DATE: 2000-02-11
; PRIOR APPLICATION NUMBER: US 60/133,200
; PRIOR FILING DATE: 1999-05-07
; PRIOR APPLICATION NUMBER: US 09/275,267
; PRIOR FILING DATE: 1998-03-23
; PRIOR APPLICATION NUMBER: US 60/119,917
; PRIOR FILING DATE: 1999-02-12
; NUMBER OF SEQ ID NOS: 1304
; SOFTWARE: Patent.pm
; SEQ ID NO 651
; LENGTH: 20674
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: 1123..3123
; OTHER INFORMATION: 5'regulatory region
; NAME/KEY: exon

```

OTHER INFORMATION:	10-513-250	: polymorphic base A or
NAME/KEY: allele		
LOCATION: 2844		
OTHER INFORMATION:	10-513-262	: polymorphic base C or T
NAME/KEY: allele		
LOCATION: 2934		
OTHER INFORMATION:	10-513-352	: polymorphic base A or G
NAME/KEY: allele		
LOCATION: 2947		
OTHER INFORMATION:	10-513-365	: polymorphic base A or G
NAME/KEY: allele		
LOCATION: 3802		
OTHER INFORMATION:	12-206-81	: polymorphic base A or G
NAME/KEY: allele		
LOCATION: 4062		
OTHER INFORMATION:	10-343-231	: deletion of C
NAME/KEY: allele		
LOCATION: 4088		
OTHER INFORMATION:	12-206-366	: polymorphic base C or T
NAME/KEY: allele		
LOCATION: 4109		
OTHER INFORMATION:	10-343-278	: polymorphic base C or T
NAME/KEY: allele		
LOCATION: 4170		
OTHER INFORMATION:	10-343-339	: polymorphic base G or T
NAME/KEY: allele		
LOCATION: 5903		
OTHER INFORMATION:	10-346-23	: polymorphic base A or G
NAME/KEY: allele		
LOCATION: 6019		
OTHER INFORMATION:	10-346-141	: polymorphic base A or G
NAME/KEY: allele		
LOCATION: 6141		
OTHER INFORMATION:	10-346-263	: polymorphic base G or C
NAME/KEY: allele		
LOCATION: 6183		
OTHER INFORMATION:	10-346-305	: polymorphic base C or T
NAME/KEY: allele		
LOCATION: 6338		
OTHER INFORMATION:	10-347-74	: polymorphic base A or G
NAME/KEY: allele		
LOCATION: 6375		
OTHER INFORMATION:	10-347-111	: polymorphic base G or C
NAME/KEY: allele		
LOCATION: 6429		
OTHER INFORMATION:	10-347-165	: polymorphic base C or T
NAME/KEY: allele		
LOCATION: 6467		
OTHER INFORMATION:	10-347-203	: polymorphic base A or G
NAME/KEY: allele		
LOCATION: 6484		
OTHER INFORMATION:	10-347-220	: polymorphic base A or G
NAME/KEY: allele		
LOCATION: 6534		
OTHER INFORMATION:	10-347-271	: polymorphic base A or T
NAME/KEY: allele		
LOCATION: 6611		
OTHER INFORMATION:	10-347-348	: polymorphic base A or G
NAME/KEY: allele		
LOCATION: 7668		
OTHER INFORMATION:	10-348-391	: polymorphic base A or G
NAME/KEY: allele		
LOCATION: 8608		
OTHER INFORMATION:	10-349-47	: polymorphic base C or T
NAME/KEY: allele		
LOCATION: 8658		
OTHER INFORMATION:	10-349-97	: polymorphic base A or G
NAME/KEY: allele		
LOCATION: 8703		
OTHER INFORMATION:	10-349-142	: polymorphic base G or C
NAME/KEY: allele		
LOCATION: 8777		
OTHER INFORMATION:	10-349-216	: deletion of CTG

[illegible]

```

RESULT 9
US-07-867-106-4/c
; Sequence 4, Application US/07867106
; Patent No. 5389526
;
; GENERAL INFORMATION:
;   APPLICANT: Slade, Martin B
;   APPLICANT: Chang, Andy C M
;   APPLICANT: Williams, Keith L
;   TITLE OF INVENTION: Improved Plasmid Vectors for Cellular
;   TITLE OF INVENTION: Slime Moulds of the Genus Dictyostelium
;   NUMBER OF SEQUENCES: 19
;
; CORRESPONDENCE ADDRESS:
;   ADDRESSEE: Woodcock Washburn Kurtz Mackiewicz & No. 5389526:ris
;   STREET: One Liberty Place 46th Floor
;   CITY: Philadelphia
;   STATE: PA
;   COUNTRY: USA
;
; ZIP: 19103
;
; COMPUTER READABLE FORM:
;   MEDIUM TYPE: Floppy disk

```



```
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA: US/07/935,313
; FILING DATE:
; APPLICATION NUMBER: EP 91 114 300.6
; FILING DATE: 26-AUG-1991
; ATTORNEY/AGENT INFORMATION:
; NAME: BENT, Stephen A.
; REGISTRATION NUMBER: 29,768
; REFERENCE/DOCKET NUMBER: 30472/114 IMMU
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (703)836-9300
; TELEFAX: (703)683-4109
; TELEX: 899149
; INFORMATION FOR SEQ ID NO: 14:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 7218 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; IMMEDIATE SOURCE:
; CLONE: pTZgpt-F1s
US-08-232-463-14

Query Match      4.0%; Score 44; DB 1; Length 7218;
Best Local Similarity 7.8%; Pred. No. 0.35;
Matches 20; Conservative 138; Mismatches 98; Indels 0; Gaps 0;

QY 99 CAAGAAATACAGATAGAAACTATTTGTTGCGAATGGAAGTAGTAAATACATTAAAG 158
||||| | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 1466 CAAGTAGTAAAGATAGAGAAATTTGGTACRRRRRRRRRRRRRRRRRRRRRRRRR 1407

QY 159 CAAATTTTAAATAATATATAGCCTATACGCGCTCAAAGTAGTATCTAGTAGTGTA 218
::: ::::: : : : : : : : : : : : : : : : : : : : : : : : : :
Db 1406 RRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRR 1347

QY 219 ATTAATAATGCGATGCGATCAGAAATGGGACAAATGAAACGGAATTTAAATATT 278
::: ::: : : : : : : : : : : : : : : : : : : : : : : : :
Db 1346 RRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRR 1287

QY 279 AACTTTAAATAATATAAATTTAGTAAATGTGTTTCTGACTATTGAGGGGCAAAAA 338
::: ::::: : : : : : : : : : : : : : : : : : : : : : : : :
Db 1286 RRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRR 1227

QY 339 AAGCAATGCCAAAAG 354
::: : : : : : : : :
Db 1226 RRRRRRRRRRRRRR 1211
```

```
RESULT 12
US-08-213-419B-3/c
; Sequence 3, Application US/08213419B
; Patent No. 6333406
; GENERAL INFORMATION:
; APPLICANT: Inselburg, J. et al.
; TITLE OF INVENTION: GENE ENCODING PROTEIN ANTIGENS OF PLASMIDIUM FALCIPARUM
; FILE REFERENCE: JII-002CNC
; CURRENT APPLICATION NUMBER: US/08/213,419B
; PRIOR FILING DATE: 1994-03-14
; PRIOR APPLICATION NUMBER: US 07/870,506
; PRIOR FILING DATE: 1992-04-17
; NUMBER OF SEQ ID NOS: 20
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 3
; LENGTH: 6124
; TYPE: DNA
; ORGANISM: Plasmodium falciparum
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (2407)..(2439)
; NAME/KEY: CDS
; LOCATION: (2598)..(3404)
```

```
; NAME/KEY: CDS
; LOCATION: (3580)..(3720)
; NAME/KEY: CDS
; LOCATION: (3850)..(5835)
US-08-213-419B-3

Query Match      4.0%; Score 43.8; DB 4; Length 6124;
Best Local Similarity 48.2%; Pred. No. 0.37;
Matches 123; Conservative 0; Mismatches 132; Indels 0; Gaps 0;

QY 720 ACAATTGATTATAGATATATCCATTAATCCATCATATTTATGATATAAATAGCTGTAAA 779
||||| | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 1884 AGAATTAATTAATTAATATTTCTAAATATATTAATTTGTAAGATACATATATATATAT 1825

QY 780 CTATTTTCAGCATCGCAGCTTTCTGCAACTTTTGTGTTTAAATTTAAAGATTTTAATAATA 839
||||| | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 1824 ATATATACATATTTGATCATGTGTACATTTTTTTTAAAGATAACAGGGATTTAGTTATCG 1765

QY 840 AAGTATTAAGAGGAGCATTAACGAGGCAACAAAGTAATGAACACGGAGAAACAAAGCCA 899
||||| | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 1764 TGACACAAAAAATATTTAAACCGGAAAAAATAAATTTAAAAAATTGTGATCTCATA 1705

QY 900 TGAAGCTCAATGTTAGTTTAAGCTTAATAAGAGATTTTAAATTTTAATGACGATG 959
||||| | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 1704 AATATATTTATTCATTCGGTTAAAAAATAAATAAATAAATAAATAAATAAATAAATA 1645

QY 960 ATAACAATTATATTT 974
||||| | | | | | | | | |
Db 1644 TATATAATAAATTT 1630

RESULT 13
US-07-867-106-2
; Sequence 2, Application US/07867106
; Patent No. 5389526
; GENERAL INFORMATION:
; APPLICANT: Slade, Martin B
; APPLICANT: Chang, Andy C M
; APPLICANT: Williams, Keith L
; TITLE OF INVENTION: Improved Plasmid Vectors for Cellular
; TITLE OF INVENTION: Slime Moulds of the Genus Dictyostellium
; NUMBER OF SEQUENCES: 19
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Woodcock Washburn Kurtz Mackiewicz & No. 5389526rls
; STREET: One Liberty Place 46th Floor
; CITY: Philadelphia
; STATE: PA
; COUNTRY: USA
; ZIP: 19103
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; FILING DATE: 19920625
; APPLICATION NUMBER: US/07/867,106
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: AU PJ 7187
; FILING DATE: 02-NOV-1989
; ATTORNEY/AGENT INFORMATION:
; NAME: Feeney, Joanne Longo
; REGISTRATION NUMBER: 35,134
; REFERENCE/DOCKET NUMBER: RICE-0002
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 215-568-3100
; TELEFAX: 215-568-3439
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 5852 base pairs
; TYPE: NUCLEIC ACID
; STRANDEDNESS: single
```

TOPOLOGY: linear
MOLECULE TYPE: DNA (genomic)
ANTI-SENSE: NO
FEATURE: CDS
NAME/KEY: CDS
LOCATION: 2378..5038
FEATURE: CDS
NAME/KEY: CDS
LOCATION: 2378..5038
US-07-867-106-2

Query Match
Best Local Similarity 4.0%; Score 43.6; DB 1; Length 5852;
Matches 121; Conservative 0; Mismatches 129; Indels 0; Gaps 0;

QY 728 TTATGATATATCCATTAATCCATGATATTTATGATATATAATAGCTGTAACTATTCA 787
DB 5586 TTATACATTTTATTTTATTTTATTTTATTTTATTTTATTTTATTTTATTTTATTTT 5645
QY 788 GCATCGCAGCTTCTGCACTTTTGTGTTTAAATTAAGAGTTTAAATTAATAAAAGTATTA 847
DB 5646 TTTTATTTTATTTTATTTTAAATTTAATTTAATTTTATTAATAAAATAGAAAT 5705
QY 848 AAGAGCATACAGGCAACAAAGTAATGAACACGAGAAACAAAGCCATGAGCTC 907
DB 5706 ATAAGTAAACAAACAAACAAATTAACATATATAAAATTAACAAATTAATAATTA 5765
QY 908 ATTGGTGTAGTTTAAAGTAAATGAAGATTTTATTAATTTTAAATGAGGATGAACAAT 967
DB 5766 AATAAATTAATACCATTAATTAATGATATATAATTAATTAATTAAGATCTTTATAGTAC 5825
QY 968 TATATTTCT 977
DB 5826 TATGTTACT 5835

RESULT 14
US-07-991-867B-32
Sequence 32, Application US/07991867B
Patent No. 5476781
GENERAL INFORMATION:
APPLICANT: Moyer, Richard W.
APPLICANT: Hall, Richard L.
APPLICANT: Gruidl, Michael E.
TITLE OF INVENTION: No. 5476781el Entomopoxvirus Expression System
NUMBER OF SEQUENCES: 66
CORRESPONDENCE ADDRESS:
ADDRESSEE: David R. Saliwanchik
STREET: 2421 N.W. 41st Street, Suite A-1
CITY: Gainesville
STATE: FL
COUNTRY: USA
ZIP: 32606
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/07/991,867B
FILING DATE: 12-DEC-1992
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: WO 92/14818
FILING DATE: 12-FEB-1992
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/827,685
FILING DATE: 30-JAN-1992
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/657,584
FILING DATE: 19-FEB-1991
ATTORNEY/AGENT INFORMATION:
NAME: Saliwanchik, David R.

REGISTRATION NUMBER: 31,794
REFERENCE/DOCKET NUMBER: UF114.C3
TELEPHONE: 904-375-8100
TELEFAX: 904-372-5800
INFORMATION FOR SEQ ID NO: 32:
SEQUENCE CHARACTERISTICS:
LENGTH: 660 base pairs
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: unknown
MOLECULE TYPE: DNA (genomic)
US-07-991-867B-32

Query Match
Best Local Similarity 3.9%; Score 43.2; DB 1; Length 660;
Matches 174; Conservative 0; Mismatches 218; Indels 0; Gaps 0;

QY 692 ATGAAGTTATAGTACTATTAGCCAGAGACAAATGATTATATAGATATATCCATTAATCCAT 751
DB 88 ATTAATATATATATATAAATAAATAGTTAATTTAGAAAGAAATTCATATATATATATAT 147
QY 752 GATATTTATGATATAAATAGCTGTAACTATTTCAGCATCGCAGCTTCTGCACTTTT 811
DB 148 GATAATAATATTAAATAATATCCAGAAATATATAAAGTTTATATATTTCAATTTA 207
QY 812 GTTTTAAATTAAGAGTTTAAATAAATAAAGTATTAATAAGGAGCATACGAGGCAACAAA 871
DB 208 AATATATTATTTAAATTTTATACAAATTAATAAATAATATACATATTTAGATATATCT 267
QY 872 AGTAATGAACACGAGGAAACAAAGCCATGAGCTCATTTGGTTAGTTTAAAGCTTAAGAAG 931
DB 268 TATAACAAAATAGCAATATAAGTAATATTATACCTACCACATCTCTATAGAATTTTAAAT 327
QY 932 AAGATTTTAAATTTTAAATGACGATGATCAATATATATTTCTGACTCTTTTAAAC 991
DB 328 TGTGAATCATGATATATAATGACTATAATTTTATTAATAATTTAGTAAATTTTAAACAAA 387
QY 992 CCCCTCTTACAAACAGAGCTCCCTTTTTCAGTAGAAGTCGAGTCCCAATCTTAAAGAC 1051
DB 388 TTAATAATATCTAAATAAATTTGGTAACTTTAATAATGTTTTCCTTAGTATAGTACTT 447
QY 1052 AAAGCCATTGAAAGAGAAAGTGAGTGAGAGA 1083
DB 448 GAGTAAATATGGAATCAATACAAATAAAGA 479

RESULT 15
US-08-107-755A-32
Sequence 32, Application US/08107755A
Patent No. 5721352
GENERAL INFORMATION:
APPLICANT: Moyer, Richard W.
APPLICANT: Hall, Richard L.
APPLICANT: Gruidl, Michael E.
TITLE OF INVENTION: No. 5721352el Entomopoxvirus Expression System
NUMBER OF SEQUENCES: 40
CORRESPONDENCE ADDRESS:
ADDRESSEE: David R. Saliwanchik
STREET: 2421 N.W. 41st Street, Suite A-1
CITY: Gainesville
STATE: Florida
COUNTRY: U.S.A.
ZIP: 32606
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/107,755A
FILING DATE: 19-AUG-1993
CLASSIFICATION: 435

;; PRIOR APPLICATION DATA:
;; APPLICATION NUMBER: US 07/827,658
;; FILING DATE: 30-JAN-1992
;; PRIOR APPLICATION DATA:
;; APPLICATION NUMBER: US 07/657,584
;; FILING DATE: 19-FEB-1991
;; ATTORNEY/AGENT INFORMATION:
;; NAME: Saliwanchik, David R.
;; REGISTRATION NUMBER: 31,794
;; REFERENCE/DOCKET NUMBER: UFI114.C2
;; TELECOMMUNICATION INFORMATION:
;; TELEPHONE: (904) 375-8100
;; TELEFAX: (904) 372-5800
;; INFORMATION FOR SEQ ID NO: 32:
;; SEQUENCE CHARACTERISTICS:
;; LENGTH: 660 base pairs
;; TYPE: nucleic acid
;; STRANDEDNESS: double
;; TOPOLOGY: unknown
;; MOLECULE TYPE: DNA (genomic)
US-08-107-755A-32

Query Match 3.9%; Score 43.2; DB 1; Length 660;
Best Local Similarity 44.4%; Pred. No. 0.29;
Matches 174; Conservative 0; Mismatches 218; Indels 0; Gaps 0;

QY	692	ATGAAAGTTATAGTACTATATAGCCAGACAGCAATTCATTATAGATATATCCATTATCCAT	751
Db	88	ATTAAT	147
QY	752	GATATTTATGATATAAATAGCTGTTAAACTATTTTCAGCATCGCAGCTTTCTGCAACTTTT	811
Db	148	GATAATAATATTTTAAATAATATCCAGAAAATATTTAAAGTTTATATATTTTCAAATTTA	207
QY	812	GTITTTTAAATTTAAGAGTTTAAATAAAGTATTTAAAGGAGCATACGAGGCAACAA	871
Db	208	AATATTTATTAATTTTAAATTTTATATAACAAATTTAAACATATTTAGATATATCT	267
QY	872	AGTAATGAACACGGAGAAACAAAGCCATGAGCTCATTTGGTTAGTTTAAAGCTTAAATAG	931
Db	268	TATACAAAATAGCAATTAAGTAATATATATCTACCACATCTATAGAAATTTTAAAT	327
QY	932	AAGATTTTATTAATTTTAAATGACGATGATACAAATATATTTCTGACTTCTTTTAAAC	991
Db	328	TGTGAATCATGTATATATAATGACTATAATTTTATTAATTAATTTAGTAAATTTAAAAAA	387
QY	992	CCCTCTTACAAACAGAGCTCCCTTTTTCAGTAGAGTCCGATCCCAATCTTAAAGAC	1051
Db	388	TAAATAATATCTAAATAATTAATTTGGTAACTTTAAATAATGTTTTCCTATTAGTATAGT	447
QY	1052	AAAGCCATTAGAAAGAGAAAGTGAGTGAGAGA	1083
Db	448	GAGTTAATATGGAATCAATACAAATAAAGA	479

Search completed: June 11, 2003, 01:53:47
Job time : 54.4398 secs

```

GenCore version 5.1.6
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run On:      June 11, 2003, 01:50:19 ; Search time 158.028 Seconds
              (without alignments)
              9750.738 Million cell updates/sec

Title:      US-09-502-426B-1_COPY_2102_3202
Perfect score: 1101
Sequence:    1 aattcacaattattaatat.....gagagagagaactagctcc 1101

Scoring table:  IDENTITY_NUC
                  Gapop 10.0, Gapext 1.0

```

```

Searched:      870385 seqs, 699768693 residues
Total number of hits satisfying chosen parameters: 1740770
  Minimum DB seq length: 0
  Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
                  Maximum Match 100%
                  Listing first 45 summaries

```

```

Database : Published_Applications_NA:*
1: /cgn2_6/ptodata/1/pubpna/US07_PUBCOMB.seq:*
2: /cgn2_6/ptodata/1/pubpna/PCF_NEW_PUB.seq:*
3: /cgn2_6/ptodata/1/pubpna/US06_NEW_PUB.seq:*
4: /cgn2_6/ptodata/1/pubpna/US06_PUBCOMB.seq:*
5: /cgn2_6/ptodata/1/pubpna/US07_NEW_PUB.seq:*
6: /cgn2_6/ptodata/1/pubpna/PCFUS_PUBCOMB.seq:*
7: /cgn2_6/ptodata/1/pubpna/US08_NEW_PUB.seq:*
8: /cgn2_6/ptodata/1/pubpna/US08_PUBCOMB.seq:*
9: /cgn2_6/ptodata/1/pubpna/US09_NEW_PUB.seq:*
10: /cgn2_6/ptodata/1/pubpna/US09_PUBCOMB.seq:*
11: /cgn2_6/ptodata/1/pubpna/US10_NEW_PUB.seq:*
12: /cgn2_6/ptodata/1/pubpna/US10_PUBCOMB.seq:*
13: /cgn2_6/ptodata/1/pubpna/US60_NEW_PUB.seq:*
14: /cgn2_6/ptodata/1/pubpna/US60_PUBCOMB.seq:*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution

```

Result No.	Score	Query Match	Length	DB ID	SUMMARIES	
					ID	Description
C 1	54.4	4.9	516	10	US-09-960-352-5785	Sequence 5785, Ap
C 2	53.2	4.8	431	10	US-09-960-352-5558	Sequence 5558, Ap
C 3	52.2	4.7	377	10	US-09-960-352-7419	Sequence 7419, Ap
C 4	52.2	4.7	446	10	US-09-960-352-3400	Sequence 3400, Ap
C 5	51.2	4.7	5690	9	US-10-239-676-73	Sequence 73, Appl
C 6	51	4.6	10286	9	US-10-239-676-14	Sequence 14, Appl
C 7	50.6	4.6	12405	9	US-10-239-676-35	Sequence 35, Appl
C 8	50.4	4.6	5689	9	US-10-239-676-90	Sequence 90, Appl
C 9	49.2	4.5	6265	10	US-09-129-112-3	Sequence 3, Appl
C 10	49.2	4.5	17419	9	US-10-239-676-100	Sequence 100, Appl
C 11	48.6	4.4	53332	9	US-10-224-562-3	Sequence 3, Appl
C 12	48.6	4.4	53332	10	US-09-801-861-3	Sequence 3, Appl
C 13	48.4	4.4	406	10	US-09-960-352-10265	Sequence 10265, A
C 14	48	4.4	525	9	US-10-198-846-1483	Sequence 1483, Ap
C 15	48	4.4	9539	9	US-10-239-676-51	Sequence 51, Appl
C 16	47.8	4.3	294	10	US-09-960-352-4637	Sequence 4637, Ap
C 17	47.8	4.3	960	9	US-10-198-846-6381	Sequence 6381, Ap
C 18	47	4.3	419	10	US-09-960-352-11234	Sequence 11234, A
C 19	46.8	4.3	6544	9	US-10-239-676-46	Sequence 46, Appl

C 20	46.6	4.2	11036	9	US-10-239-676-118	Sequence 118, Appl
C 21	46.6	4.2	17848	9	US-10-239-676-28	Sequence 28, Appl
C 22	46.4	4.2	364	9	US-09-191-995-18370	Sequence 18370, A
C 23	46.4	4.2	960	9	US-10-198-846-6381	Sequence 6381, Ap
C 24	46.2	4.2	1713	9	US-09-938-842A-4756	Sequence 4756, Ap
C 25	46.2	4.2	7823	9	US-10-239-676-158	Sequence 158, Appl
C 26	46	4.2	12405	9	US-10-239-676-36	Sequence 36, Appl
C 27	46	4.2	15732	9	US-10-239-676-96	Sequence 96, Appl
C 28	46	4.2	32463	9	US-09-990-936-5	Sequence 5, Appli
C 29	45.8	4.2	428	10	US-09-960-352-573	Sequence 573, App
C 30	45.8	4.2	8996	9	US-10-239-676-211	Sequence 211, Appl
C 31	45.8	4.2	9539	9	US-10-239-676-52	Sequence 52, Appl
C 32	45.6	4.1	1109	9	US-09-938-842A-3333	Sequence 3333, Ap
C 33	45.6	4.1	7153	9	US-10-239-676-70	Sequence 70, Appl
C 34	45.6	4.1	640681	10	US-09-790-988-1	Sequence 1, Appli
C 35	45.2	4.1	6204	9	US-10-239-676-190	Sequence 190, App
C 36	45	4.1	413	10	US-09-960-352-2919	Sequence 2919, App
C 37	45	4.1	5917	9	US-10-087-464-9	Sequence 9, Appli
C 38	45	4.1	7195	9	US-10-239-676-30	Sequence 30, Appl
C 39	45	4.1	7657	9	US-10-239-676-185	Sequence 185, App
C 40	44.6	4.1	6306	9	US-10-239-676-224	Sequence 224, App
C 41	44.6	4.1	7903	9	US-10-239-676-110	Sequence 110, App
C 42	44.4	4.0	4985	9	US-10-094-240-10	Sequence 10, Appl
C 43	44.4	4.0	6397	9	US-10-239-676-107	Sequence 107, Appl
C 44	44.2	4.0	344	10	US-09-960-352-1036	Sequence 1036, Ap
C 45	44.2	4.0	4187	9	US-10-072-349-252	Sequence 252, App

ALIGNMENTS

RESULT 1

US-09-960-352-5785/c

; Sequence 5785, Application US/09960352

; Patent No. US20020137139A1

; GENERAL INFORMATION:

; APPLICANT: Warren, Wesley C.

; APPLICANT: Tao, Nengbing

; APPLICANT: Byatt, John C.

; APPLICANT: Mathialagan, Nagappan

; TITLE OF INVENTION: NUCLEIC ACID AND OTHER MOLECULES ASSOCIATED WITH LACTATION AND

; TITLE OF INVENTION: NUCLEIC AND FAT DEPOSITION

; FILE REFERENCE: 16511.006/37-21(10298)C

; CURRENT APPLICATION NUMBER: US/09/960.352

; CURRENT FILING DATE: 2001-09-24

; NUMBER OF SEQ ID NOS: 15112

; SEQ ID NO 5785

; LENGTH: 516

; TYPE: DNA

; ORGANISM: Bos taurus

; FEATURE:

; NAME/KEY: unsure

; LOCATION: (76),(90)

; OTHER INFORMATION: unsure at all n locations

; OTHER INFORMATION: Clone ID: 25-LIB3057-018-Q1-K1-G1

IS-09-960-352-5785

	Query Match	4.9%;	Score 54.4;	DB 10;	Length 516;
	Best Local Similarity	48.7%;	Pred. No. 0.14;	Mismatches 0;	Gaps 0;
	Matches 148;	Conservative			
Y	5	TACAAATTTATATATTTTACTCAATACAAATGCATGCAAGTTCCAAAAAAATTTTGT	64		
b	460	TAAAAAAATATATAATAAAAAATAAAAATAAAAAAATAAAAATATAA	401		
Y	65	AACAGAACTCCAAATTTTTTTTTTATGGACACGAAGAAATACAGATAGAACTATT	124		
b	400	AAAAAAATAAAAATATAAAAAATATAAAAAAAATAAAAATATAAATAAAAACAAAA	341		
Y	125	TTGTTGTGGAATGGAAGTAGTAAATACATTAAACAAATTTTAAAAAAATATATAAGCCT	184		
b	340	TTATAAAAAAATAAAAATAAAAAAATAAAAAAATAAAAATAAAAAAATAAAAATAAAA	281		

TYPE: DNA
ORGANISM: Dolichos biflorus
FEATURE:
OTHER INFORMATION: genomic sequence of NBP46 (DB46)
NAME/KEY: exon
LOCATION: (633)..(944)
NAME/KEY: intron
LOCATION: (945)..(1022)
NAME/KEY: exon
LOCATION: (1023)..(1151)
NAME/KEY: intron
LOCATION: (1152)..(1559)
NAME/KEY: exon
LOCATION: (1560)..(1616)
NAME/KEY: intron
LOCATION: (1617)..(1697)
NAME/KEY: exon
LOCATION: (1698)..(1790)
US-09-129-112-3

Query Match
Best Local Similarity 4.5%; Score 49.2; DB 10; Length 6265;
Matches 173; Conservative 0; Mismatches 148; Indels 5; Gaps 3;
QY 15 TAATATTAGTCATACATGCGATAGAAAGTTCCAAAAAATTTTGTAAACAGAAACT 74
DB 2936 TATTTATAGTGTCTTACTATTCAATAAACAATAATGCGAGATATGAAGTTTACTG-AACT 2878
QY 75 TCCAAATTTTTTTTTTATGGAACAAGAAATAACAGATAGAAACTATTTTGTGTGA 134
DB 2877 TCTTTTTTTTTTTTTTTTATGGAATGTTTGAATATGTCGAATATTAATATAGTGA 2818
QY 135 ATGGAAGTAGTAATACATTTAA--GCAATTTTAAAAATTTAATAAGCCTATACGCGC 192
DB 2817 ATATTATAGAAATATGCAATAATTTGAAACTGAAAAAATAATTAATTTAATAGTGA 2758
QY 193 TCAAAGTAGTTTATCTAGTAGTGTAATTAATA--ATGCAATGTCGCGATTTCAGAAATGGG 250
DB 2757 TCACAACAAGTTTGAATTAATTTAAACTATATTAATTAATTAATTAATTAATTAATTAAT 2698
QY 251 ACAACAATGAACGGAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 310
DB 2697 CTATAAATAAATAAATAATTTAATAAATAAATAAATAAATAAATAAATAAATAAATAAATA 2638
QY 311 TGTTCCTGACTATTGAGGGCAGAA 336
DB 2637 TTTTATATTCAATTAAGCAGACAA 2612

RESULT 10
US-10-239-676-100/c
Sequence 100, Application US/10239676
Publication No. US20030082609A1
GENERAL INFORMATION:
APPLICANT: OLEK, Alexander
APPLICANT: PIEPENBROCK, Christian
APPLICANT: BERLIN, Kurt
TITLE OF INVENTION: Diagnosis of Diseases Associated with Gene Regulation
FILE REFERENCE: 5013.1003
CURRENT APPLICATION NUMBER: US/10/239,676
CURRENT FILING DATE: 2002-09-24
PRIOR APPLICATION NUMBER: PCT/EP01/03968
DE 10019058.8
DE 10019173.8
DE 10032529.7
DE 10043826.1
PRIOR FILING DATE: 2001-04-06
2000-04-06
2000-04-07
2000-06-30
2000-09-01
NUMBER OF SEQ ID NOS: 228
SEQ ID NO 100

LENGTH: 17419
TYPE: DNA
ORGANISM: Artificial Sequence
FEATURE:
OTHER INFORMATION: chemically treated genomic DNA (Homo sapiens)
US-10-239-676-100
Query Match
Best Local Similarity 4.5%; Score 49.2; DB 9; Length 17419;
Matches 173; Conservative 0; Mismatches 188; Indels 1; Gaps 1;
QY 660 ATCTAATGACATTTTTTTTTTCTACACGCGTGGATGAAAGTTATATCTACTATATGACGAG 719
DB 11067 ATCAATATCTTACTTACTTAAATCTTTTCAACATATATCTTACTTAACTATATTTCTACT 11008
QY 720 ACAATTGATATAGATATATCCATTAATCCATGATATTTATGATATAAATAGCTGTGTTAA 779
DB 11007 TTACTTCTTATTTTATTTTACTTATTAATCAAACTTACTTCTTTTCCAAATACATTTTAT 10948
QY 780 CTATTTACGATCGCAGCTTTCTGCAACTTTTGTGTTTAAATTAAGATTTTAATAATAA 839
DB 10947 TTTAATAATAAAATCTTCTTTTAAAAAACTAATTAATAAATAAATTAATAATAAATAA 10888
QY 840 AAGTATTAAGGAGCATACGAGCGACAAAGTAATGAACGAGAGAAACAAAGCCA 899
DB 10887 TAATAATACAATAATTTAATACAATAATATACATCTAAACCAATACTAAATAATAA 10828
QY 900 TGAAGCTCATTTGGTTAGTTTAAGCTTAATAGAGATTTTATTAATTTTAATGACGATG 959
DB 10827 CTCA-CATAATATCTTTCGTTAATTAATAAACAATAAATAATATATTTATCAACTCATA 10769
QY 960 ATACAATTAATTTTCTGACTTCTTTAAACCCCTCTTTACAAACAGAGCTCCCTTTT 1019
DB 10768 AAAACAATACATATATATCTTATTTTCTATATCTTATCTATATCTTAAACCTTAAACCCCTATT 10709
QY 1020 TC 1021
DB 10708 TC 10707

RESULT 11
US-10-224-562-3/c
Sequence 3, Application US/10224562
Publication No. US2003002229A1
GENERAL INFORMATION:
APPLICANT: YAN, Chunhua et al.
TITLE OF INVENTION: ISOLATED HUMAN KINASE PROTEINS, NUCLEIC
TITLE OF INVENTION: ACID MOLECULES ENCODING HUMAN KINASE PROTEINS, AND USES
FILE REFERENCE: CL001098DIV
CURRENT APPLICATION NUMBER: US/10/224,562
CURRENT FILING DATE: 2002-08-21
NUMBER OF SEQ ID NOS: 10
SOFTWARE: FastSeq for Windows Version 4.0
SEQ ID NO 3
LENGTH: 53332
TYPE: DNA
ORGANISM: Homo sapiens
US-10-224-562-3
Query Match
Best Local Similarity 4.4%; Score 48.6; DB 9; Length 53332;
Matches 129; Conservative 0; Mismatches 134; Indels 0; Gaps 0;
QY 47 TCCAAAAAAATTTTGTAAACAGAACTTCCAAATTTTCTTTTATGGAACAGAAAT 106
DB 31671 TCAAAAAAAATTAATACATATATATATAAAAAAATATATAATATAATAATATAT 31612
QY 107 AACAGATAGAAACTATTTTGTGGAATGGAAGTACTAATATACATTAAGCAAAATTT 166
DB 31611 AAATATATAAATATATATATATAATAATAATAATAATAATAATAATAATAATA 31552
QY 167 AAAAAATTAATAGCCTATACGCGCTCAAAAGTATGTATCTAGTAGTGTAAATAA 226


```
Db      88 AAAAAAAAAACANNAGANANANAAAAAAAAAAAAAAAAAAAAAAAAAAAAAC 147
QY      61 TGTACAGAACTCCAAATTTTTTTTTTATGAGACAGAAATACGATGAAAC 120
Db      148 ANAAAAAANCANNANANANANANANANANANANANANANANANANANANAN 207
QY      121 TATTTGTTGTAAGTGAAGTACTAATATACATTAGCAATTTTAAAAATATATA 180
Db      208 AANNANANANANANANANANANANANANANANANANANANANANANANANAN 267
QY      181 GCCTACGCCCTAAAGTATGTTATCTAGTAGGTAAATTAATGATGTCGATT 240
Db      268 NAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAN 327
QY      241 CAGATTGGACAACAATGAAACGCAATTAATTAATTAATTAATTAATTAATTA 300
Db      328 AAAAAAAAAANANANANANANANANANANANANANANANANANANANANANAN 387
QY      301 TGAGTAAATGTGTTTCTGACTATGAGGGCAAAAAAAAAAGACATGCCAAAGCTA 358
Db      388 TNAATAAATATNTAANANANANANANANANANANANANANANANANANANAT 445
```

RESULT 15

```
US-10-239-676-51/C
; Sequence 51, Application US/10239676
; Publication No. US2003082609A1
; GENERAL INFORMATION:
; APPLICANT: OLEK, Alexander
; APPLICANT: PIEPENBROCK, Christian
; APPLICANT: BERLIN, Kurt
; TITLE OF INVENTION: Diagnosis of Diseases Associated with Gene Regulation
; FILE REFERENCE: 5013.1003
; CURRENT APPLICATION NUMBER: US/10/239,676
; CURRENT FILING DATE: 2002-09-24
; PRIOR APPLICATION NUMBER: PCT/EP01/03968
; DE 10019058.8
; DE 10019173.8
; DE 10032529.7
; DE 10043826.1
; PRIOR FILING DATE: 2001-04-06
; 2000-04-06
; 2000-04-07
; 2000-06-30
; 2000-09-01
; NUMBER OF SEQ ID NOS: 228
; SEQ ID NO 51
; LENGTH: 9539
; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: chemically treated genomic DNA (Homo sapiens)
US-10-239-676-51
```

Query Match 4.4%; Score 48; DB 9; Length 9539;
Best Local Similarity 47.9%; Pred. No. 7.9;

Matches 138; Conservative 0; Mismatches 150; Indels 0; Gaps 0;

```
QY      27 AATAACAATGCATAGAAAGTCCAAAAAATTTTGTAAACAGAACTCCAAATTTT 86
Db      6023 ATTAAAAAATAATATAAATCAATTCATAAAATCAAAAAATCTAAAAAATCAA 5964
QY      87 TTTTGTGGAACAAGAAATACAGATGAAACTATTTGTTGTGAATGGAAGTAGTA 146
Db      5963 AATTTCTAAAAAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 5904
QY      147 ATATACATTAGCAATTTTAAAAAATTAATTAAGCCTATACGCGCTCAAGTAGTAT 206
Db      5903 TAACATAATAAAAAAATTTAAAAACCAAAATTAATTAATTAATTAATTAATTA 5844
QY      207 CTAGTAGTGTAATTAATTAATGCAATGCGATTCAGAAATGGACAAATGAAAAACG 266
```

```
Db      5843 CTAATAAATAATTCAAAAAACAATAAATATATTTTAAACCTAAATACGATTTAAAAA 5784
QY      267 AATTAAATATTAACCTTTAAATTAATTAATTAATTAATTAATTAATTAATTAAT 314
Db      5783 ACAAAACATATAAATTTCAATACGACAAACAAATTTTAAAAAATTAACCTT 5736
```

Search completed: June 11, 2003, 08:07:37
Job time: 159.028 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: June 10, 2003, 22:03:39 ; Search time 2645.14 Seconds
(without alignments)
10465.184 Million cell updates/sec

Title: US-09-502-426B-1_COPY_2102_3202

Perfect score: 1101
Sequence: 1 aatcacaataattataat.....gagagagagaactagctcc 1101

Scoring table:
IDENTITY NUC
Gapop 10.0 , Gapext 1.0

Searched: 24791104 segs, 12571243825 residues

Total number of hits satisfying chosen parameters: 49582208

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%

Listing first 45 summaries

Database :

Pending_Patents_MN_Main:*

1: /cgn2_6/ptodata/1/pna/US06000.COMB.seq.*
2: /cgn2_6/ptodata/1/pna/US06001.COMB.seq.*
3: /cgn2_6/ptodata/1/pna/US06002.COMB.seq.*
4: /cgn2_6/ptodata/1/pna/US06003.COMB.seq.*
5: /cgn2_6/ptodata/1/pna/US06004.COMB.seq.*
6: /cgn2_6/ptodata/1/pna/US06005.COMB.seq.*
7: /cgn2_6/ptodata/1/pna/US06006.COMB.seq.*
8: /cgn2_6/ptodata/1/pna/US06007.COMB.seq.*
9: /cgn2_6/ptodata/1/pna/US06008.COMB.seq.*
10: /cgn2_6/ptodata/1/pna/US06009.COMB.seq.*
11: /cgn2_6/ptodata/1/pna/US06010.COMB.seq.*
12: /cgn2_6/ptodata/1/pna/US06011.COMB.seq.*
13: /cgn2_6/ptodata/1/pna/US06012.COMB.seq.*
14: /cgn2_6/ptodata/1/pna/US06013.COMB.seq.*
15: /cgn2_6/ptodata/1/pna/US06014.COMB.seq.*
16: /cgn2_6/ptodata/1/pna/US06015.COMB.seq.*
17: /cgn2_6/ptodata/1/pna/US06016.COMB.seq.*
18: /cgn2_6/ptodata/1/pna/US06017.COMB.seq.*
19: /cgn2_6/ptodata/1/pna/US06018.COMB.seq.*
20: /cgn2_6/ptodata/1/pna/US06019.COMB.seq.*
21: /cgn2_6/ptodata/1/pna/US06020.COMB.seq.*
22: /cgn2_6/ptodata/1/pna/US06021.COMB.seq.*
23: /cgn2_6/ptodata/1/pna/US06022.COMB.seq.*
24: /cgn2_6/ptodata/1/pna/US06023.COMB.seq.*
25: /cgn2_6/ptodata/1/pna/US06024.COMB.seq.*
26: /cgn2_6/ptodata/1/pna/US06025.COMB.seq.*
27: /cgn2_6/ptodata/1/pna/US06026.COMB.seq.*
28: /cgn2_6/ptodata/1/pna/US06027.COMB.seq.*
29: /cgn2_6/ptodata/1/pna/US06028.COMB.seq.*
30: /cgn2_6/ptodata/1/pna/US06029.COMB.seq.*
31: /cgn2_6/ptodata/1/pna/US06030.COMB.seq.*
32: /cgn2_6/ptodata/1/pna/US06031.COMB.seq.*
33: /cgn2_6/ptodata/1/pna/US06032.COMB.seq.*
34: /cgn2_6/ptodata/1/pna/US06033.COMB.seq.*
35: /cgn2_6/ptodata/1/pna/US06034.COMB.seq.*
36: /cgn2_6/ptodata/1/pna/US06035.COMB.seq.*
37: /cgn2_6/ptodata/1/pna/US06036.COMB.seq.*
38: /cgn2_6/ptodata/1/pna/US06037.COMB.seq.*
39: /cgn2_6/ptodata/1/pna/US06038.COMB.seq.*
40: /cgn2_6/ptodata/1/pna/US06039.COMB.seq.*
41: /cgn2_6/ptodata/1/pna/US06040.COMB.seq.*
42: /cgn2_6/ptodata/1/pna/US06041.COMB.seq.*
43: /cgn2_6/ptodata/1/pna/US06042.COMB.seq.*

Result No.	Score	Query Match	Length	DB ID	Description
1	1101	100.0	6888	1	Sequence 1, Appl1
2	1101	100.0	6888	1	Sequence 1, Appl1
3	1090	99.0	84196	27	Sequence 94, Appl1
4	1090	99.0	84196	31	Sequence 1074, Ap
5	960.4	87.2	1699	18	Sequence 2050, Ap
6	125	11.4	826	1	Sequence 3486, Ap
7	66.6	6.0	469	33	Sequence 40543, A
8	66.6	6.0	469	33	Sequence 84816, A
9	65	5.9	450	17	Sequence 15000, A
10	65	5.9	528	25	Sequence 121875, A
11	65	5.9	528	27	Sequence 121875, A
12	61	5.5	553	22	Sequence 29073, A
13	60.4	5.5	507	33	Sequence 91952, A
14	60.4	5.5	507	64	Sequence 135967, A
15	59.8	5.4	358	25	Sequence 340019, A
16	59.8	5.4	358	27	Sequence 340019, A
17	59.8	5.4	358	58	Sequence 6397, Ap
18	59.6	5.4	422	33	Sequence 30413, Ap
19	59.6	5.4	547	33	Sequence 1402, Ap
20	59.6	5.4	547	64	Sequence 1402, Ap
21	5.4	5.4	664	29	Sequence 739, App

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

22	59	5.4	664	31	US-09-803-110-739	Sequence 739, App
c 23	58.8	5.3	2000	33	US-09-887-7728-5563	Sequence 5263, App
c 24	58.4	5.3	447	25	US-09-654-617-341141	Sequence 341141, App
c 25	58.4	5.3	447	27	US-09-664-016-341141	Sequence 341141, App
c 26	58.4	5.3	447	58	US-60-145-485-9030	Sequence 9030, App
c 27	58.4	5.3	720	18-	US-09-404-520-12953	Sequence 12953, App
c 28	58.4	5.3	805	56	US-60-126-265-101-127	Sequence 127, App
c 29	58.4	5.3	961	56	US-60-126-265-3039	Sequence 3039, App
c 30	57.6	5.2	513	25	US-09-654-617-181496	Sequence 181496, App
c 31	57.6	5.2	513	27	US-09-664-016-181496	Sequence 181496, App
c 32	57.2	5.2	389	33	US-09-885-4194-17337	Sequence 17337, App
c 33	57.2	5.2	676	33	US-09-885-4194-7909	Sequence 7909, App
c 34	56.8	5.2	485	23	US-09-619-643-59	Sequence 59, App
c 35	56.8	5.2	485	25	US-09-654-617-334306	Sequence 334306, App
c 36	56.8	5.2	485	27	US-09-664-016-334306	Sequence 334306, App
c 37	56.8	5.1	485	58	US-60-146-907-71	Sequence 71, App
c 38	56.6	5.1	407	24	US-09-637-0864-28965	Sequence 28965, App
c 39	56.6	5.1	407	24	US-09-637-0864-28965	Sequence 28965, App
c 40	56.6	5.1	6107	42	US-10-221-613-304	Sequence 304, App
c 41	56.4	5.1	469	25	US-09-654-617-86668	Sequence 86668, App
c 42	56.4	5.1	469	27	US-09-664-016-86668	Sequence 86668, App
c 43	56.2	5.1	534	22	US-09-565-306-13731	Sequence 13731, App
c 44	56	5.1	501	33	US-09-873-4024-2082	Sequence 2082, App
c 45	56	5.1	501	64	US-60-209-830-2082	Sequence 2082, App

ALIGNMENTS

RESULT 1
 PCT-US00-03820-1
 : Sequence 1, Application PC/TUS00003820
 : GENERAL INFORMATION:
 : APPLICANT: THE ARIZONA BOARD OF REGENTS ON BEHALF OF THE UNIVERSITY OF ARIZONA
 : TITLE OF INVENTION: DMF4 POLYNUCLEOTIDES, POLYPEPTIDES AND USES THEREOF
 : FILE REFERENCE: 2225-0001.40
 : CURRENT APPLICATION NUMBER: PCT/US00/03820
 : CURRENT FILING DATE: 2000-02-11
 : EARLIER APPLICATION NUMBER: 60/119,657
 : EARLIER FILING DATE: 1999-02-11
 : EARLIER APPLICATION NUMBER: 60/119,658
 : EARLIER FILING DATE: 1999-02-11
 : NUMBER OF SEQ ID NOS: 18
 : SOFTWARE: PatentIn Ver. 2.0
 : SEQ ID NO 1
 : LENGTH: 6888
 : TYPE: DNA
 : ORGANISM: Arabidopsis sp.
 PCT-US00-03820-1

Query Match	100.0%	Score 1101	DB 1	Length 6888
Best Local Similarity	100.0%	Pred. No. 2,4e-189		
Matches 1101	Conservative 0	Mismatches 0	Indels 0	Gaps 0
QY	1	AATCTACAAATATTAATATTAGTCATCAATGATAGAAAATGCCAAAAAAATTT	60	
Db	2102	AATCTACAAATATTAAATATTAGTCATCAATGATAGAAAATGCCAAAAAAATTT	2166	
QY	61	TGTTAACAGAACTCCAAATTTTTTTTTTATAGAACAGAAATAACAGATAGAAAC	120	
Db	2162	TGTTAACAGAACTCCAAATTTTTTTTTTATAGAACAGAAATAACAGATAGAAAC	2222	
QY	121	TATTTTGTGTGGAATGGAAGTAGTAATATACATTAGCAAAATTTTAAAAAATTATATTA	180	
Db	2222	TATTTTGTGTGGAATGGAAGTAGTAATATACATTAGCAAAATTTTAAAAAATTATATTA	2281	
QY	181	GCCATACGCCCTCAAAAGTAGTATCTAGAGGGTAAATTAATATATGCAATGGCCGATT	240	
Db	2282	GCCATACGCCCTCAAAAGTAGTATCTAGAGGGTAAATTAATATATGCAATGGCCGATT	2341	
QY	241	CAGAAATGGGCAACAATGAAGAAAGGAATTTAAATTAATTAATTAATTAATAAATTT	300	
Db	2342	CAGAAATGGGCAACAATGAAGAAAGGAATTTAAATTAATTAATTAATTAATAAATTT	2401	

OY	301	TGAGTAAATGTTTTCTGACTATTGAGGGCAAAAAAGCAATGCGAAAGCTACG	360
Db	2402	TGAGTAAATGTTTTCTGACTATTGAGGGCAAAAAAGCAATGCGAAAGCTACG	2465
OY	361	GGTTTGACTGTCAGTTGGTAAATATCTAAATCTGCTTTTGACGCGACGCTGCT	420
Db	2462	GGTTTGACTGTCAGTTGGTAAATATCTAAATCTGCTTTTGACGCGACGCTGCT	2521
OY	421	AGGGGTCCTTGACATTTTTCACGTTTTTACCCTACGCGAGCGCCACTTTTCCAT	480
Db	2522	AGGGGTCCTTGACATTTTTCACGTTTTTACCCTACGCGAGCGCCACTTTTCCAT	2581
OY	481	ATCCTAAGGGTAATTTTGGAAATCCCAATTTAAACCGATGAGACCGTACCGACTTCT	540
Db	2582	ATCCTAAGGGTAATTTTGGAAATCCCAATTTAAACCGATGAGACCGTACCGACTTCT	2642
OY	541	GGGATTCCTGCTGAGACATTTATCAAAAATTTAGACAGAAATGGGTTATTATTTAAA	600
Db	2642	GGGATTCCTGCTGAGACATTTATCAAAAATTTAGACAGAAATGGGTTATTATTTAAA	2701
OY	601	ACTCAACAACCTGATCAGATATAAATTTCTAATAACCTTTAGATGAGATCTGATGATCTA	660
Db	2702	ACTCAACAACCTGATCAGATATAAATTTCTAATAACCTTTAGATGAGATCTGATGATCTA	2761
OY	661	TCTATGACACTTTTTTTTTTCTACACACGCTGATGAAAGTTATGACTATTAGCCAGAGA	720
Db	2762	TCTATGACACTTTTTTTTTTCTACACACGCTGATGAAAGTTATGACTATTAGCCAGAGA	2822
OY	721	CAATTTGATTTATAGATATATTCATTATNTCCATTATGATATTTATGATATTAATAGCTGTTAAAC	780
Db	2822	CAATTTGATTTATAGATATATTCATTATNTCCATTATNTGATATTTAAATTAAGCTGTTAAAC	2881
OY	781	TATTTTCAGCATCGGAGGCTTCTGCAACTTTTGTTTTTAATTTAAGAGTTTATATAATATAA	840
Db	2882	TATTTTCAGCATCGGAGGCTTCTGCAACTTTTGTTTTTAATTTAAGAGTTTATATAATATAA	2942
OY	841	AGTATTTAAAGAGCATATACGAGGCAACAAAGTATATGATGACACGAGAGCAAAAGCCAT	900
Db	2942	AGTATTTAAAGAGCATATACGAGGCAACAAAGTATATGATGACACGAGAGCAAAAGCCAT	3001
OY	901	GAACCTCATGCTGTAGTTTAGCTTAATATAGAAAGTTTTATTAATTTTATATGACAGATGA	960
Db	3002	GAACCTCATGCTGTAGTTTAGCTTAATATAGAAAGTTTTATTAATTTTATATGACAGATGA	3061
OY	961	TAAACATTTATTTTCTGACTCTTTTAAACCCCTCTTTACAAACAGAGCTCCCTTTTT	1020
Db	3062	TAAACATTTATTTTCTGACTCTTTTAAACCCCTCTTTACAAACAGAGCTCCCTTTTT	3122
OY	1021	CAGTAGAAGTCCGATTTCCCAATCTTTAAAGCAAAAGCCATTGAAAGACAAAGTCAGTAG	1080
Db	3122	CAGTAGAAGTCCGATTTCCCAATCTTTAAAGCAAAAGCCATTGAAAGACAAAGTCAGTAG	3181
OY	1081	AGAGAGAGAGAACTAGCTCC 1101	
Db	3182	AGAGAGAGAGAACTAGCTCC 3202	

RESULT 2
 US-09-502-426-1
 : Sequence 1, Application US/09502426
 :
 : GENERAL INFORMATION:
 :
 : APPLICANT: AZPIROZ, Ricardo
 :
 : APPLICANT: CHOE, Sunghwa
 :
 : APPLICANT: FELDMANN, Kenneth
 :
 : TITLE OF INVENTION: DMP4 POLYNUCLEOTIDES, POLYPEPTIDES AND USES THEREOF
 :
 : FILE REFERENCE: 2225-0001
 :
 : CURRENT APPLICATION NUMBER: US/09/502,426
 :
 : CURRENT FILING DATE: 2000-02-11
 :
 : EARLIER APPLICATION NUMBER: 60/119,657
 :
 : EARLIER FILING DATE: 1999-02-11
 :
 : EARLIER APPLICATION NUMBER: 60/119,658
 :
 : EARLIER FILING DATE: 1999-02-11

NUMBER OF SEQ ID NOS: 18
SOFTWARE: Patentin Ver. 2.0
SEQ ID NO 1
LENGTH: 6888
TYPE: DNA
ORGANISM: Arabidopsis sp.
US-09-502-426-1

Query Match 100.0%; Score 1101; DB 19; Length 6888;
Best Local Similarity 100.0%; Pred. No. 2.4e-189;
Matches 1101; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

1 AATCTACAAATTTATTTATTTAGTCAATTAACATGATAGAGAAAGTCCAAAAAATTT 60
2102 AATCTACAAATTTATTTATTTAGTCAATTAACATGATAGAGAAAGTCCAAAAAATTT 2161
61 TGTTAACAGAACTCCAAATTTTTTTTTTATGAAACAGAAATTAACAGATGAAAC 120
2162 TGTTAACAGAACTCCAAATTTTTTTTTTATGAAACAGAAATTAACAGATGAAAC 2221
121 TATTTGTTGGAAATGAGTAAATATATCATATGAAATTTTAAATTTATTTA 180
2222 TATTTGTTGGAAATGAGTAAATATATCATATGAAATTTTAAATTTATTTA 2281
181 GCTTATAGCGCTCAAGATATGTAATCTAGTGTATTAATTAATGATGTCGAT 240
2282 GCTTATAGCGCTCAAGATATGTAATCTAGTGTATTAATTAATGATGTCGAT 2341
241 CAGAAATGGGACACAAATGAAACGAAATTAATTAATTAATTAATTAATTAAT 300
2342 CAGAAATGGGACACAAATGAAACGAAATTAATTAATTAATTAATTAATTAAT 2401
301 TGAGTAATATGTTTCTGACATTTGAGGGGCAAAAAAGCAATGCAAAAGTCTAC 360
2402 TGAGTAATATGTTTCTGACATTTGAGGGGCAAAAAAGCAATGCAAAAGTCTAC 2461
361 GGTGACATGTCAGTGGTAAATATATCAATCACTGCTTTGACCGCAGCTGCT 420
2462 GGTGACATGTCAGTGGTAAATATATCAATCACTGCTTTGACCGCAGCTGCT 2521
421 AGGGTCTCTGACATTTTCACTGCTTACCCCTACCTGACGCGCCACCTTTCCAT 480
2522 AGGGTCTCTGACATTTTCACTGCTTACCCCTACCTGACGCGCCACCTTTCCAT 2581
481 ATCTAAGGTAATTTTGGAAATGCCAATTAACCGATGAGCGTACCGGACTGCT 540
2582 ATCTAAGGTAATTTTGGAAATGCCAATTAACCGATGAGCGTACCGGACTGCT 2641
541 GGGATTCGCTGAGCATTTATCAAAATTTATGACAGCAATGGGTTTATTAATTA 600
2642 GGGATTCGCTGAGCATTTATCAAAATTTATGACAGCAATGGGTTTATTAATTA 2701
601 ACTCAACATTTATCAATTAATTAATTAATTAATTAATTAATTAATTAATTA 660
2702 ACTCAACATTTATCAATTAATTAATTAATTAATTAATTAATTAATTAATTA 2761
661 TCTAATGACTTTTTTTCTACACGGTGAAGTAAAGTAAATGATCTTTAGCCAG 720
2762 TCTAATGACTTTTTTTCTACACGGTGAAGTAAAGTAAATGATCTTTAGCCAG 2821
721 CAATGATTAATGATATATCATTAATCATGATTTATGATATTAATAGCTGTTAAC 780
2822 CAATGATTAATGATATATCATTAATCATGATTTATGATATTAATAGCTGTTAAC 2881
781 TATTTACGATCGCAGCTTTGCAACTTTGTTTTTAAATTAATTAATTAATTA 840
2882 TATTTACGATCGCAGCTTTGCAACTTTGTTTTTAAATTAATTAATTAATTA 2941
841 ACTAATTAAGGATTAAGGAGCAAAAGTAAATGAAACGAGGAAACAAAGCCAT 900
2942 ACTAATTAAGGATTAAGGAGCAAAAGTAAATGAAACGAGGAAACAAAGCCAT 3001
901 GAAGTCAATGGTATGTTAAGCTTAATTAAGAAATTTTAAATTTAATGACGATGA 960

|||||
Db 3002 GAAGTCAATGGTATGTTAAGCTTAATTAAGAAATTTTAAATTTAATGACGATGA 3061
961 TACAAATTAATTTTCTGACTTTTAAACCCCTTTACAAACAGAGCTCCCTTTT 1020
3062 TACAAATTAATTTTCTGACTTTTAAACCCCTTTACAAACAGAGCTCCCTTTT 3121
1021 CAGTGAATGCGATTCCTCAATTAAGAGCAACGCTTGAAGAGAGAGTGTAG 1080
3122 CAGTGAATGCGATTCCTCAATTAAGAGCAACGCTTGAAGAGAGAGTGTAG 3181
1081 AGAGAGAGAGAACTAGCTCC 1101
3182 AGAGAGAGAGAACTAGCTCC 3202

RESULT 3
US-09-692-412-94/C
Sequence 94, Application US/09692412
GENERAL INFORMATION:
APPLICANT: Bush, David F.
APPLICANT: Rounsley, Steven D.
APPLICANT: Wiegand, Roger C.
TITLE OF INVENTION: Plant Polymorphic Markers and Uses Thereof
FILE REFERENCE: 38-10(15493)C
CURRENT APPLICATION NUMBER: US/09/692,412
CURRENT FILING DATE: 2000-10-24
NUMBER OF SEQ ID NOS: 124
SEQ ID NO 94
LENGTH: 84196
TYPE: DNA
ORGANISM: Arabidopsis thaliana
US-09-692-412-94

Query Match 99.0%; Score 1090; DB 27; Length 84196;
Best Local Similarity 99.9%; Pred. No. 3.6e-187;
Matches 1101; Conservative 0; Mismatches 0; Indels 1; Gaps 1;
1 AATCTACAAATTTATTTATTTAGTCAATTAACATGATAGAGAAAGTCCAAAAAATTT 60
30101 AATCTACAAATTTATTTATTTAGTCAATTAACATGATAGAGAAAGTCCAAAAAATTT 30042
61 TGTTAACAGAACTCCAAATTTTTTTTTTATGAAACAGAAATTAACAGATGAAAC 120
30041 TGTTAACAGAACTCCAAATTTTTTTTTTATGAAACAGAAATTAACAGATGAAAC 29982
121 TATTTGTTGGAAATGAGTAAATATATCATATGAAATTTTAAATTTATTA 180
29981 TATTTGTTGGAAATGAGTAAATATATCATATGAAATTTTAAATTTATTA 29922
181 GCTTATAGCGCTCAAGATATGTAATCTAGTGTATTAATTAATGATGTCGAT 240
29921 GCTTATAGCGCTCAAGATATGTAATCTAGTGTATTAATTAATGATGTCGAT 29862
241 CAGAAATGGGACACAAATGAAACGAAATTAATTAATTAATTAATTAATTAAT 300
29861 CAGAAATGGGACACAAATGAAACGAAATTAATTAATTAATTAATTAATTAAT 29802
301 TGAGTAATATGTTTCTGACATTTGAGGGGCAAAAAAGCAATGCAAAAGTCTAC 360
29801 TGAGTAATATGTTTCTGACATTTGAGGGGCAAAAAAGCAATGCAAAAGTCTAC 29742
361 GGTGACATGTCAGTGGTAAATATATCAATCACTGCTTTGACCGCAGCTGCT 420
29741 GGTGACATGTCAGTGGTAAATATATCAATCACTGCTTTGACCGCAGCTGCT 29682
421 AGGGTCTCTGACATTTTCACTGCTTACCCCTACTGCTGAGCCACCTTTCCAT 480
29681 AGGGTCTCTGACATTTTCACTGCTTACCCCTACTGCTGAGCCACCTTTCCAT 29622
481 ATCTAAGGTAATTTTGGAAATGCCAATTAATTAACCGATGAGCGGATGCTGCT 540
29621 ATCTAAGGTAATTTTGGAAATGCCAATTAATTAACCGATGAGCGGATGCTGCT 29562

QY	541	GGGATTCGCGGAGCATTATACAAAATTTATGACAGCAATGGGTTATTAATTAATAA	600
Db	29561	GGGATTCGCGGAGCATTATACAAAATTTATGACAGCAATGGGTTATTAATTAATAA	29562
QY	601	ACTCAACAATTGATCAGATATAAATTTTCATAAACCTTTTACGATGATTCGTACATCTA	660
Db	29501	ACTCAACAATTGATCAGATATAAATTTTCATAAACCTTTTACGATGATTCGTACATCTA	29442
QY	661	TCTATATGCTTTTTTTTTTTTCTACACAGGTGATGAAGTATATGACTATTAGCCAGAGA	720
Db	29441	TCTATATGCTTTTTTTTTTTTCTACACAGGTGATGAAGTATATGACTATTAGCCAGAGA	29382
QY	721	CAATTGATTTATAGATATATATCCATTATATCCATGATATTTATGATATTAATATAGCTGTTAAAC	780
Db	29381	CAATTGATTTATAGATATATATCCATTATATCCATGATATTTATGATATTAATATAGCTGTTAAAC	29322
QY	781	TATTTGACATCCGACGCTTTGTGCACCTTTGTTTTTATTTAAGAGCTTTAATAATATAA	840
Db	29321	TATTTGACATCCGACGCTTTGTGCACCTTTGTTTTTATTTAAGAGCTTTAATAATATAA	29262
QY	841	AGATTTAAAGAGACATATACGAGCAACAAAGATATATACACGAGAGAAACAAAGCCAT	900
Db	29261	AGATTTAAAGAGACATATACGAGCAACAAAGATATATACACGAGAGAAACAAAGCCAT	29202
QY	901	GAACCTCATTTGGTTA-GTTTTAGCTTATATAAGAGATTTTATTAATTTTATGACGATG	959
Db	29201	GAACCTCATTTGGTTAGTTTAAAGCTTATATAAGAGATTTTATTAATTTTATGACGATG	29142
QY	960	ATAACAAATTAATTTTCTGACTCTTTTAAACCCCTCTTCAACAACAGACCTCCCTTTT	1019
Db	29141	ATAACAAATTAATTTTCTGACTCTTTTAAACCCCTCTTCAACAACAGACCTCCCTTTT	29082
QY	1020	TCAGTAGAAGTCGAGTCCCAATCTTAAAGACAAGCCATAGAGAAAGAAAGTAGATGA	1079
Db	29081	TCAGTAGAAGTCGAGTCCCAATCTTAAAGACAAGCCATAGAGAAAGAAAGTAGATGA	29022
QY	1080	GAGAGAGAGAGAAACTAGCTCC	1101
Db	29021	GAGAGAGAGAGAAACTAGCTCC	29000
RESULT 4			
US-09-803-736-1074/c			
Sequence 1074, Application US/09803736			
GENERAL INFORMATION:			
APPLICANT: Bush, David F.			
APPLICANT: Levin, Irene M.			
APPLICANT: Norris, Susan R.			
APPLICANT: Rounsley, Steven D.			
APPLICANT: Wiegand, Roger C.			
TITLE OF INVENTION: Plant Polymorphic Markers and Uses Thereof			
FILE REFERENCE: 38-10(115493)D			
CURRENT APPLICATION NUMBER: US/09/803, 736			
CURRENT FILING DATE: 2001-03-12			
PRIOR APPLICATION NUMBER: US 09/534, 859			
PRIOR FILING DATE: 2000-03-29			
PRIOR APPLICATION NUMBER: Identified by Attorney Docket number 04983.0206CPUS01			
PRIOR FILING DATE: 2000-10-20			
NUMBER OF SEQ ID NOS: 1582			
SEQ ID NO 1074			
LENGTH: 84196			
TYPE: DNA			
ORGANISM: Arabidopsis thaliana			
US-09-803-736-1074			

Query Match	99.0%;	Score 1090;	DB 31	Length 84156;
Best Local Similarity	99.9%;	Pred. No. 3 6e-187;		
Matches 1101;	Conservative 0;	Mismatches 0;	Indels 1;	Gaps 1;
OY	1	AAATCTACAAATTTTAAATATTGTCATATACATGCATAGAAAGTTCACAAAAAATTTT	60	
Db	30101	AAATCTACAAATTTTAAATATTGTCATATACATGCATAGAAAGTTCACAAAAAATTTT	3004.2	

OY	61	GGTTAAGCAAGAACTCCAAATTTTTTTTTTTTNGSAAACAAGAAATPAACGATGAGAAAC	120
Db	30041	TGTTACAGAAAGCTCCAAATTTTTTTTTTTTNGSAAACAAGAAATPAACGATGAGAAAC	29988
OY	121	TATTTTGTGTGGAAGTGAAGTACTAATATACATTAAGCAAAATTTTAAAAATTTATATNA	180
Db	29981	TATTTTGTGTGGAAGTGAAGTACTAATATACATTAAGCAAAATTTTAAAAATTTATATNA	29922
OY	181	GCGATATCGGCGCTCAAGTAGTATGTAATGACTAGTGAATTAATATGATGGTGGATT	240
Db	29921	GCGATATCGGCGCTCAAGTAGTATGTAATGACTAGTGAATTAATATGATGGTGGATT	29867
OY	241	CAGAAATTTGGGCAACAATGAAAAAGCAATTAATAATTAATTAATTAATTAATTAATTAAT	300
Db	29861	CAGAAATTTGGGCAACAATGAAAAAGCAATTAATAATTAATTAATTAATTAATTAATTAAT	29802
OY	301	TGACTAAATGTGTTTTCTGACTATTGAGGGGCAAAAAAGACAAATGCCAAAGTACG	360
Db	29801	TGACTAAATGTGTTTTCTGACTATTGAGGGGCAAAAAAGACAAATGCCAAAGTACG	29742
OY	361	GGTTTGACTGCTCCAGTTGCGGTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT	420
Db	29741	GGTTTGACTGCTCCAGTTGCGGTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT	29688
OY	421	AGGGGTCCTTGACATTTTCACTGTTCTACCCCTACTCGTGAGCCACCCCTTTCCAT	480
Db	29681	AGGGGTCCTTGACATTTTCACTGTTCTACCCCTACTCGTGAGCCACCCCTTTCCAT	29622
OY	481	ATCCTAAGGGTAATTTTGGAAATCCCAATTAACGATGAGAACCGTACCGGACTCCCT	540
Db	29621	ATCCTAAGGGTAATTTTGGAAATCCCAATTAACGATGAGAACCGTACCGGACTCCCT	29562
OY	541	GGGATTTCTGCTGAGCACTTATCAAAAATTTATTAGCAGCAATGGCTTTATTAATTAAAA	600
Db	29561	GGGATTTCTGCTGAGCACTTATCAAAAATTTATTAGCAGCAATGGCTTTATTAATTAAAA	29502
OY	601	ACTACACACTTGATCAATATAAATTTATATAACCTTTTACGATGGATTGCTACGATCTA	660
Db	29501	ACTACACACTTGATCAATATAAATTTATATAACCTTTTACGATGGATTGCTACGATCTA	29442
OY	661	TCTAATACCTTTTTTTTTTTTACACACGGTGGATGAAGTAATGTACTATTAGCCAGAGA	720
Db	29441	TCTAATACCTTTTTTTTTTTTACACACGGTGGATGAAGTAATGTACTATTAGCCAGAGA	29382
OY	721	CAATTGATATAGATATATCCATTATATCATGATATTTATGATATATAATAGCTGTTAAAC	780
Db	29381	CAATTGATATAGATATATCCATTATATCATGATATTTATGATATATAATAGCTGTTAAAC	29322
OY	781	TATTTGAGCAATCGAGCTTCTGCAACTTTGTTTTTAATTAAGGTTTAATTAATTAATAA	840
Db	29321	TATTTGAGCAATCGAGCTTCTGCAACTTTGTTTTTAATTAAGGTTTAATTAATTAATAA	29262
OY	841	AGTATTTAAAAAGGACATACAGGCAACAACAAGTATGAAACAAGGAAAAACAAGCCAT	900
Db	29261	AGTATTTAAAAAGGACATACAGGCAACAACAAGTATGAAACAAGGAAAAACAAGCCAT	29202
OY	901	GAAAGCTATAGTGTTA-GTTTAAGCTTATTAAGAAAGATTTTATTAATTTTATGACGATG	959
Db	29201	GAAAGCTATAGTGTTAAGCTTATTAAGAAAGATTTTATTAATTTTATGACGATG	29142
OY	960	ATAACAAATTAATTTTCTGACTTCTTTAAAAACCCCTCTTACAAACAGAGCTCCCTTTT	1019
Db	29141	ATAACAAATTAATTTTCTGACTTCTTTAAAAACCCCTCTTACAAACAGAGCTCCCTTTT	29082
OY	1020	TCAGTAGAGCTCGGATTTCCCAATCTTAAAGACAAGCCATTAAGAAAGAGAAAGTAGTGA	1079
Db	29081	TCAGTAGAGCTCGGATTTCCCAATCTTAAAGACAAGCCATTAAGAAAGAGAAAGTAGTGA	29022
OY	1080	GAGAGAGAGAGAAACTAGCTCC 1101	
Db	29021	GAGAGAGAGAGAAACTAGCTCC 29000	

RESULT 5
 US-09-451-320-2050
 ; Sequence 2050, Application US/09451320
 ; GENERAL INFORMATION:
 ; APPLICANT: ALEXANDROV, Nikolai
 ; APPLICANT: CHEN, Xianfeng
 ; APPLICANT: SUBRAMANIAN, Gopalakrishnan
 ; APPLICANT: ZHENG, Liansheng
 ; TITLE OF INVENTION: SEQUENCE-DETERMINED DNA FRAGMENTS 42-59, 62-66, 68, 81
 ; FILE REFERENCE: 2750-0662P
 ; CURRENT APPLICATION NUMBER: US/09/451,320
 ; NUMBER OF SEQ ID NOS: 6998
 ; SOFTWARE: PatentIn Ver. 2.1
 ; SEQ ID NO 2050
 ; LENGTH: 1699
 ; TYPE: DNA
 ; ORGANISM: Arabidopsis thaliana
 ; FEATURE:
 ; OTHER INFORMATION: any n or Xaa = unknown
 US-09-451-320-2050

Query Match
 Best Local Similarity 87.2%; Score 960.4; DB 18; Length 1699;
 Matches 983; Conservative 0; Mismatches 12; Indels 2; Gaps 2;

QY 107 AACAGATAGAAAACCTATTTTGTGGAATGGAAGTAGTAATATCATTAAGCAAAATTTT 166
 DB 1 AACAGATAGAAAACCTATTTTGTGGAATGGAAGTAGTAATATCATTAAGCAAAATTTT 166
 QY 167 AAAAAATTTATTAAGCCTATACGCCCTCAAAAGTATGTATCTAGTAGGTATTAATAA 226
 DB 61 AAAAAATTTATTAAGCCTATACGCCCTCAAAAGTATGTATCTAGTAGGTATTAATAA 226
 QY 227 TGCATGTGCGATTTAGAAATGGGACAAATGAAAGAAATTAATTAATTAATA 120
 DB 121 TGCATGTGCGATTTAGAAATGGGACAAATGAAAGAAATTAATTAATTAATA 120
 QY 287 AATAATTAATAATTTAGTAATATGTTTCTGACTTGTAGGGGCAAAAAAGACAAT 180
 DB 181 AATAATTAATAATTTAGTAATATGTTTCTGACTTGTAGGGGCAAAAAAGACAAT 180
 QY 347 GCCAAAAGTCTACGGGTTTGTGCTGCTCAGTTCGTAATATCATTAATCATCTGCTTGA 240
 DB 241 GCCAAAAGTCTACGGGTTTGTGCTGCTCAGTTCGTAATATCATTAATCATCTGCTTGA 240
 QY 407 CCGCAGCTGCTGTAGGGGTCCTTCTGACATTTTCTGCTTACCCCTACTCGTAGCC 300
 DB 301 CCGCAGCTGCTGTAGGGGTCCTTCTGACATTTTCTGCTTACCCCTACTCGTAGCC 300
 QY 467 CACCCCTTTCCCATATCTTAAGGTAATTTTGAATCCCAATTTAAACCGATTGAGACC 360
 DB 361 CACCCCTTTCCCATATCTTAAGGTAATTTTGAATCCCAATTTAAACCGATTGAGACC 360
 QY 527 GTCAGGACTTCCCTGGATTCGTCGAGCATTTATCAAAAATTAATTAAGCAGATGGT 420
 DB 421 GTCAGGACTTCCCTGGATTCGTCGAGCATTTATCAAAAATTAATTAAGCAGATGGT 420
 QY 587 TTAATTAATTAATAAATCTACAACCTGATGATTAATTTTCAATTAACACTTTAGCATGG 480
 DB 481 TTAATTAATTAATAAATCTACAACCTGATGATTAATTTTCAATTAACACTTTAGCATGG 480
 QY 647 ATTCTAGCATCTATTAATGACTTTTCTTCTACGAGGATGGAAGATTAATGA 540
 DB 541 ATTCTAGCATCTATTAATGACTTTTCTTCTACGAGGATGGAAGATTAATGA 540
 QY 707 CTATTAGCAGAGACAATGATTAATGATTAATCATTAATCCATGATTAATTAATGA 600
 DB 601 CTATTAGCAGAGACAATGATTAATGATTAATCATTAATCCATGATTAATTAATGA 600
 QY 767 AATAGCTGTAACCTTTTACAGATGCGAGCTTTCGCAACTTTGTTTAATTAATA 826
 DB 826 AATAGCTGTAACCTTTTACAGATGCGAGCTTTCGCAACTTTGTTTAATTAATA 826

DB 661 AATAGCTGTAACCTTTTACAGATGCGAGCTTTCGCAACTTTGTTTAATTAATA 720
 QY 827 GTTTAATTAATAAAGTATTAATAAGSACATTAAGAGCAACAAAGTAATGAACAGCA 886
 DB 721 GTTTAATTAATAAAGTATTAATAAGSACATTAAGAGCAACAAAGTAATGAACAGCA 780
 QY 887 GAAACAAAAGCCATGACCTATTTGTTA-GTTTAAGCTTAATTAAGAAATTTAATAA 945
 DB 781 GAAACAAAAGCCATGACCTATTTGTTAAGCTTAATTAAGAAATTTAATAA 945
 QY 946 TTTTAATGAC-GATGATTAACATTAATTTTCTGACTTCTTTAAACCCCTTACAAA 840
 DB 841 TTTTAATGACGATGATTAACATTAATTTTCTGACTTCTTTAAACCCCTTACAAA 900
 QY 1005 CAGAGCTCCCTTTTCTAGTAAGATTCGATTCCTCAATCTTAAGCAACGATTAATA 1064
 DB 901 CAGAGCTCCCTTTTCTAGTAAGATTCGATTCCTCAATCTTAAGCAACGATTAATA 1064
 QY 1065 AGAGAAAGTGTAGAGAGAGAGAGAGAGAACTAGCTCC 1101
 DB 961 AGAGAAAGTGTAGAGAGAGAGAGAGAGAACTAGCTCC 997

RESULT 6
 PCT-US99-22853B-3486
 ; Sequence 3486, Application PC/TUS9922853B
 ; GENERAL INFORMATION:
 ; APPLICANT: Ceres, Inc.
 ; TITLE OF INVENTION: Sequence-Determined DNA Fragments and Corresponding
 ; FILE REFERENCE: Polypeptides Encoded Thereby
 ; CURRENT APPLICATION NUMBER: PCT/US99/22853B
 ; NUMBER OF SEQ ID NOS: 3938
 ; SOFTWARE: MS Word 97
 ; SEQ ID NO 3486
 ; LENGTH: 826
 ; TYPE: DNA
 ; ORGANISM: Arabidopsis thaliana
 ; FEATURE:
 ; OTHER INFORMATION: LOCATION 1..826, Ceres Seq. ID 1592085
 ; LOCATION: (1)..(826)
 ; OTHER INFORMATION: any n = any nucleic acid, unknown or other
 PCT-US99-22853B-3486

Query Match
 Best Local Similarity 11.4%; Score 125; DB 1; Length 826;
 Matches 125; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 977 TGACTTCTTAAACCCCTCTTACAAACAGAGCTCCCTTTTCTAGTAGAAGTCGAT 1036
 DB 1 TGACTTCTTAAACCCCTCTTACAAACAGAGCTCCCTTTTCTAGTAGAAGTCGAT 60
 QY 1037 CCCAATCTTAAGACAAAGCCATTAAGAAAGAAAGTGAAGAGAGAGAGAAACTA 1096
 DB 61 CCCAATCTTAAGACAAAGCCATTAAGAAAGAAAGTGAAGAGAGAGAGAAACTA 120
 QY 1097 GCTCC 1101
 DB 121 GCTCC 125

RESULT 7
 US-09-865-439A-40543/C
 ; Sequence 40543, Application US/09865439A
 ; GENERAL INFORMATION:
 ; APPLICANT: Edgerton, Michael D
 ; APPLICANT: Hardeman, Kristine J.
 ; APPLICANT: La Rosa, Thomas J.
 ; TITLE OF INVENTION: Nucleic Acid Molecules And Other Molecules Associated With
 ; FILE REFERENCE: 38-21(51936)B

```

Query Match      5.9%; Score 65; DB 17; Length 450;
Best Local Similarity 100.0%; Pred. No. 0.06;
Matches 65; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY      1037 CCCAATCTTAAAGCAAAAGCCATTAGAAAGAGAAAGTGAGTGAGAGAGAGAGAAACTA 10986
          |||||
Db       55   CCCAATCTTAAAGCAAAAGCCATTAGAAAGAGAAAGTGAGTGAGAGAGAGAGAAACTA 114

QY      1097 GCTCC 1101
          |||||
Db      115 GCTCC 119

RESULT 10
US-09-654-617-121875
; Sequence 121875, Application US/09654617
; GENERAL INFORMATION:
; APPLICANT: Kovalic, David K.
; APPLICANT: Liu, Jingdong

```

;; TITLE OF INVENTION: Annotated Plant Genes
;; FILE REFERENCE: 38-21(15097)D
;; CURRENT APPLICATION NUMBER: US/09/654,617
;; CURRENT FILING DATE: 2000-09-05
;; NUMBER OF SEQ ID NOS: 463173
;; SEQ ID NO 121875
;; LENGTH: 528
;; TYPE: DNA
;; ORGANISM: Arabidopsis thaliana
;; OTHER INFORMATION: unsure at all n locations
US-09-654-617-121875

Query Match 5.9%; Score 65; DB 25; Length 528;
Best Local Similarity 100.0%; Pred. No. 0.062;
Matches 65; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1037 CCCATCTTAAAGACAAAGCCATTGAAAGAAAGTGTGAGAGAGAGAGAACTA 1096
|||
DB 88 CCCATCTTAAAGACAAAGCCATTGAAAGAAAGTGTGAGAGAGAGAGAACTA 147

QY 1097 GCTCC 1101
|||||
DB 148 GCTCC 152

RESULT 11
US-09-684-016-121875
; Sequence 121875, Application US/09684016
; GENERAL INFORMATION:
; APPLICANT: Kovacic, David K.
; APPLICANT: Liu, Jindong
; TITLE OF INVENTION: Annotated Plant Genes
; FILE REFERENCE: 38-21(15097)D
; CURRENT APPLICATION NUMBER: US/09/684,016
; CURRENT FILING DATE: 2000-10-10
; PRIOR FILING DATE: 2000-09-05
; NUMBER OF SEQ ID NOS: 463173
; SEQ ID NO 121875
; LENGTH: 528
; TYPE: DNA
; ORGANISM: Arabidopsis thaliana
; FEATURE:
; NAME/KEY: unsure
; LOCATION: (1)..(528)
; OTHER INFORMATION: unsure at all n locations
US-09-684-016-121875

Query Match 5.9%; Score 65; DB 27; Length 528;
Best Local Similarity 100.0%; Pred. No. 0.062;
Matches 65; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1037 CCCATCTTAAAGACAAAGCCATTGAAAGAAAGTGTGAGAGAGAGAACTA 1096
|||
DB 88 CCCATCTTAAAGACAAAGCCATTGAAAGAAAGTGTGAGAGAGAGAACTA 147

QY 1097 GCTCC 1101
|||||
DB 148 GCTCC 152

RESULT 12
US-09-565-306-29073
; Sequence 29073, Application US/09565306
; GENERAL INFORMATION:
; APPLICANT: Andersen, Scott E.
; APPLICANT: Conner, Timothy W.
; APPLICANT: Lajudt, Raghuath V.
; TITLE OF INVENTION: Nucleic Acid Molecules and Other Molecules Associated with
; FILE REFERENCE: 38-21(15459)C
; CURRENT APPLICATION NUMBER: US/09/565,306
; CURRENT FILING DATE: 2000-05-04

;; NUMBER OF SEQ ID NOS: 83523
;; SEQ ID NO 29073
;; LENGTH: 553
;; TYPE: DNA
;; ORGANISM: Zea mays
;; OTHER INFORMATION: Clone ID: CC-zml189094b09d1
US-09-565-306-29073

Query Match 5.5%; Score 61; DB 22; Length 553;
Best Local Similarity 50.1%; Pred. No. 0.33;
Matches 177; Conservative 0; Mismatches 175; Indels 1; Gaps 1;

QY 1 AATCTACAAATTTATATATTAGTCATTAACATGATGAGAAAGTCCAAAAAATTT 60
|||
DB 76 AAAAAATTAATTTATATATTTAAATTTAAATTTAAATTTAAATTTAAATTTAA 135
|||
QY 61 TGTAAACAGAACTCCAAATTTTATTTTATGAGCAAGAAATTAACAGATGAAC 120
|||
DB 136 AATTAATTAATTTATTAATTTATTAATTTATTAATTTATTAATTTATTAATTT 195
|||
QY 121 TATTTTGTGTGAGATGAGATGATGATGATGATGATGATGATGATGATGATG 180
|||
DB 196 AATTAATTAATTTATTAATTTATTAATTTATTAATTTATTAATTTATTAATTT 255
|||
QY 181 GCCATACGCGCTCAAGTATGTTATCTAGTAGTGTATTAATGATGATGATGAT 240
|||
DB 256 AATTAATTAATTTATTAATTTATTAATTTATTAATTTATTAATTTATTAATTT 314
|||
QY 241 CAGATTTGGACACACATGAAACCGAATTAATTTAATTTAATTTAATTTAATTT 300
|||
DB 315 ATTAATTTAATTTAATTTAATTTAATTTAATTTAATTTAATTTAATTTAATTT 374
|||
QY 301 TGAGTAATGCTTTTCTGACTATGAGGGCAAAAAAGACATGCGCAAA 353
|||
DB 375 TAAAAAATCATTAATTAATTTAATTTAATTTAATTTAATTTAATTTAATTT 427
|||

RESULT 13
US-09-865-439A-91952/C
; Sequence 91952, Application US/09865439A
; GENERAL INFORMATION:
; APPLICANT: Edgerton, Michael D
; APPLICANT: Hardeman, Kristine J.
; APPLICANT: La Rosa, Thomas J.
; TITLE OF INVENTION: Nucleic Acid Molecules And Other Molecules Associated with
; FILE REFERENCE: 38-21(51936)B
; CURRENT APPLICATION NUMBER: US/09/865,439A
; CURRENT FILING DATE: 2001-05-29
; PRIOR FILING DATE: 2000-05-30
; NUMBER OF SEQ ID NOS: 119126
; SEQ ID NO 91952
; LENGTH: 507
; TYPE: DNA
; ORGANISM: Zea mays
; FEATURE:
; NAME/KEY: unsure
; LOCATION: (1)..(507)
; OTHER INFORMATION: unsure at all n locations
; OTHER INFORMATION: Clone ID: LIB3607-057-Q6-N6-H11
US-09-865-439A-91952

Query Match 5.5%; Score 60.4; DB 33; Length 507;
Best Local Similarity 51.7%; Pred. No. 0.42;
Matches 136; Conservative 0; Mismatches 127; Indels 0; Gaps 0;

QY 46 TTCAAAAAATTTGTTACAGAACTCCAAATTTTATTTTATGAGCAAGAA 105
|||
DB 237 TTTTAAAAAATTTTAAAAAATTTTATTTTATTTTATTTTATTTTATTTTATTT 238
|||
QY 106 TAACAGATGAAATATTTTGTGTGAGATGAGATGAGATGAGATGAGATGAGAT 165
|||

Db 237 AAAATTTTAAATATATTTTTTTTTTAAACAAAAATCTTTTTTTTTTAAACAAAAA 178
Qy 166 TAAAAATTTATTAAGCTTAACGGCTCAAGATGTTAGTACTAGCTGAATTAATA 225
Db 177 AAAAAAATTTTAAAGCTTTTAAACAAAAATCTTTTTTTTTTAAACAAAAA 118
Qy 226 ATGCATGTCGATTCAGATTTGGGACAAATGAAACGAAATTAATTAATTAATTA 285
Db 117 AAAAAAATTTTAAAGCTTTTAAACAAAAATCTTTTTTTTTTAAACAAAAA 58
Qy 286 AAATTAATTAATTTTGAATTA 308
Db 57 AAAAAAATTTTAAAGCTTTTAAACAAAAATCTTTTTTTTTTAAACAAAAA 35

RESULT 14
US-60-207-458-135967/c
; Sequence 135967, Application US/60207458
; GENERAL INFORMATION:
; APPLICANT: Abad, Mark S.
; APPLICANT: Conner, Timothy W.
; APPLICANT: Delkman, Jill
; APPLICANT: Hardeman, Kristine J.
; APPLICANT: La Rosa, Thomas J.
; APPLICANT: Lalgudi, Raghunath V.
; APPLICANT: Ruan, Yijun G.
; APPLICANT: Ruff, Thomas G.
; APPLICANT: Sammons, R. Douglas
; APPLICANT: Shukla, Hridayabhiranjan
; APPLICANT: Wu, Kunsheng
; APPLICANT: Xu, Nanfei
; TITLE OF INVENTION: NUCLEIC ACID MOLECULES AND OTHER MOLECULES ASSOCIATED WITH
; FILE REFERENCE: 38-21(51936)A
; CURRENT APPLICATION NUMBER: US/60/207,458
; CURRENT FILING DATE: 2000-05-30
; NUMBER OF SEQ ID NOS: 152403
; SEQ ID NO 135967
; LENGTH: 507
; TYPE: DNA
; ORGANISM: Zea mays
; OTHER INFORMATION: Clone ID: LIB3607-057-Q6-N6-H11
US-60-207-458-135967

Query Match 5.5%; Score 60.4; DB 64; Length 507;
Best Local Similarity 51.7%; Pred. No. 0.42; Mismatches 127; Indels 0; Gaps 0;
Matches 136; Conservative 0;

Qy 46 TTCCAAAAAATTTTGTAAACAGAACTTCCAAATTTTTTTTTTATGAGACAGAA 105
Db 297 TTTTAAATTTTAAATTTTAAACAAAAATTTTTTTTTTTTTTTTCCCATTA 238
Qy 106 TAACAGATGAAACATTTTGTGTGAATGGAATGATTAATTAATTAATTAATTA 165
Db 237 AAAAAATTTTAAATTTTAAATTTTAAACAAAAATCTTTTTTTTTTAAACAAAA 178
Qy 166 TAAAAATTTTAAAGCTTTTAAACAAAAATCTTTTTTTTTTAAACAAAAA 225
Db 177 AAAAAAATTTTAAAGCTTTTAAACAAAAATCTTTTTTTTTTAAACAAAAA 118
Qy 226 ATGCATGTCGATTCAGATTTGGGACAAATGAAACGAAATTAATTAATTAATTA 285
Db 117 AAAAAAATTTTAAAGCTTTTAAACAAAAATCTTTTTTTTTTAAACAAAAA 58
Qy 286 AAATTAATTAATTTGAGTAA 308
Db 57 AAAAAAATTTTAAAGCTTTTAAACAAAAATCTTTTTTTTTTAAACAAAAA 35

RESULT 15
US-09-654-617-340019/c
; Sequence 340019, Application US/09654617
; GENERAL INFORMATION:

; APPLICANT: Kovalic, David K.
; APPLICANT: Liu, Jindong
; TITLE OF INVENTION: Annotated Plant Genes
; FILE REFERENCE: 38-21(15097)D
; CURRENT APPLICATION NUMBER: US/09/654,617
; CURRENT FILING DATE: 2000-09-05
; NUMBER OF SEQ ID NOS: 463173
; SEQ ID NO 340019
; LENGTH: 358
; TYPE: DNA
; ORGANISM: Zea mays
US-09-654-617-340019

Query Match 5.4%; Score 59.8; DB 25; Length 358;
Best Local Similarity 49.8%; Pred. No. 0.51;
Matches 151; Conservative 0; Mismatches 152; Indels 0; Gaps 0;

Qy 6 ACAATTTTATTTTATTTAGTCAATAGATGATGAAAGTCCAAAAATTTTGTGA 65
Db 337 ATAAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 278
Qy 66 ACAGAACTTCCAAATTTTATGAAACAAATTAAGATGAAACATTAATTT 125
Db 277 AAAAAAATTTTAAATTTTAAATTTTAAACAAAAATTTTTTTTAAATTAATTA 218
Qy 126 TGTGTGGAAGGAGTAAATTAATTAATTAATTAATTAATTAATTAATTAATTA 185
Db 217 AATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 158
Qy 186 TACGGCCCAAGTATGTTATGTAATTAATTAATTAATTAATTAATTAATTAAT 245
Db 157 ATATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 98
Qy 246 TTGGCAACAATGAAACGAAATTAATTAATTAATTAATTAATTAATTAATTA 305
Db 97 AATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 38
Qy 306 AAA 308
Db 37 AAA 35

Search completed: June 11, 2003, 05:00:37
Job time : 2648.14 secs


```
QY 901 GAAGCTCATGTTAGTTAAGCTTAATAGACATTTTAAATTTTATGACATGA 960
    |||
Db 3002 GAAGCTCATGTTAGTTAAGCTTAATAGACATTTTAAATTTTATGACATGA 3061
QY 961 TAACATTTATTTTCTGACTTTTAAACCCTCTTAACAAGAGAGCTCTTTT 1020
    |||
Db 3062 TAACATTTATTTTCTGACTTTTAAACCCTCTTAACAAGAGAGCTCTTTT 3121
QY 1021 CAGTAAAGTCCGATCCCAATCTTAAGCAAGGCATTAAGAAAGAGAGTGA 1080
    |||
Db 3122 CAGTAAAGTCCGATCCCAATCTTAAGCAAGGCATTAAGAAAGAGAGTGA 3181
QY 1081 AGAGAGAGAGAACTAGCTCC 1101
    |||
Db 3182 AGAGAGAGAGAACTAGCTCC 3202
```

RESULT 3

```
US-60-466-412-85418
; Sequence 85418, Application US/60466412
; GENERAL INFORMATION:
; APPLICANT: CARGILL, Michele
; APPLICANT: TAKOHOVA, Olga
; TITLE OF INVENTION: GENETIC POLYMORPHISMS ASSOCIATED WITH
; FILE REFERENCE: CL001466
; CURRENT APPLICATION NUMBER: US/60/466,412
; CURRENT FILING DATE: 2003-04-30
; NUMBER OF SEQ ID NOS: 429241
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 85418
; LENGTH: 16885
; TYPE: DNA
; ORGANISM: Homo sapiens
US-60-466-412-85418
```

```
Query Match
Best Local Similarity 5.2%; Score 57.2; DB 10; Length 16885;
Matches 190; Conservative 0; Mismatches 203; Indels 1; Gaps 1;
```

```
QY 581 ATGGCTTATTTATTTAAACTCACAACCTGATCAGATTAATTTCAATTAACCTTTTA 640
    |||
Db 4838 ATGGCTTATTTATTTAAAGTTATTTACTTAATTAACCTCCCTTT-ATATATATATATA 4896
QY 641 CGATGATTCGTACGATCTATCTAATGACTTTTCTTCTTCTACGAGGTGAAGTT 700
    |||
Db 4897 ATATATATATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 4956
QY 701 ATAGATCTATTAAGCAGACAAATGATTATGATTAATTCATTAATCCATGATTTAT 760
    |||
Db 4957 ATATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 5016
QY 761 GATATAATACGCTGTTAAGTATTTTACATCGAGCTTCTGCACTTTTGTTTTAT 820
    |||
Db 5017 TATATAATTTATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 5076
QY 821 TTAAGACTTTTAAATTAAGTATTAAGAGACATTAAGAGCAAGCAAGTAAAGAA 880
    |||
Db 5077 TTAATATATATATATATATTAATTAATTAATTAATTAATTAATTAATTAATTA 5136
QY 881 CACGAGAGAAACAAAGCAGATGATGCTGTTAGTTAAGCTTAATTAAGAGATTTTA 940
    |||
Db 5137 ATATATTTATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 5196
QY 941 TTAATTTTAAATGAGATGATTAATTAATTAATTTT 974
    |||
Db 5197 TAAATTTATTAATTAATTAATTAATTAATTAATTAAT 5230
```

```
RESULT 4
US-09-615-606A-35136/C
; Sequence 35136, Application US/09615606A
; GENERAL INFORMATION:
```

```
; APPLICANT: Abad, Mark S.
; APPLICANT: Buehler, Robert E.
; APPLICANT: Byrum, Joseph R.
; APPLICANT: Coombs, Brian E.
; APPLICANT: Heck, Gregory R.
; APPLICANT: La Rosa, Thomas J.
; APPLICANT: Nelson, Donald E.
; APPLICANT: Shukla, Hridayabhiranjan
; APPLICANT: Thompson, Michael D.
; TITLE OF INVENTION: Nucleic Acid Molecules And Other Molecules Associated With
; FILE REFERENCE: 38-21(15444)C
; CURRENT APPLICATION NUMBER: US/09/615,606A
; CURRENT FILING DATE: 2000-07-13
; NUMBER OF SEQ ID NOS: 91663
; SEQ ID NO 35136
; LENGTH: 458
; TYPE: DNA
; ORGANISM: Glycine max
; FEATURE:
; NAME/KEY: unsure
; LOCATION: (1)..(458)
; OTHER INFORMATION: unsure at all n locations
; FEATURE:
; OTHER INFORMATION: Clone ID: LIB3053-011-01-N1-C9
US-09-615-606A-35136
```

```
Query Match
Best Local Similarity 5.0%; Score 55; DB 6; Length 458;
Matches 178; Conservative 0; Mismatches 208; Indels 0; Gaps 0;
```

```
QY 1 AATCTACAAATTTATTTATTTAGTCAATTAACATGATAGAAAGTCCAAAAAATTT 60
    |||
Db 406 AATCTATATATATATATATATATATATATATATATATATATATATATATATATAT 347
QY 61 TGTTCACAGAACTCCAAATTTTATTTTATGACACAGAAATTAACATAGAAAC 120
    |||
Db 346 AATATAAATAAATAATATATATATATATATATATATATATATATATATATATATAT 287
QY 121 TATTTTGTGTGGAATGAGAGTATATATATATATATATATATATATATATATATAT 180
    |||
Db 286 TAAAAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATA 227
QY 181 GCGTATACGCGCTCAAGTATGTTATCTAGTAGTGTTATTAATTAATGATGTCGATT 240
    |||
Db 226 ATATAGATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 167
QY 241 CAGATTTGGACACAAATGAAGGATTTAAATATTTAATTAATTAATTAATTAATTAAT 300
    |||
Db 166 AATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAA 107
QY 301 TGAATTAATGTTTCTGACTATTTGAGGGGCAAAAAAAGACAATGCAAAAGCTACG 360
    |||
Db 106 AATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAA 47
QY 361 GGTGACTGCTGCGATGCGTAATA 386
    |||
Db 46 AATTAACGATTTAATTAAGGAAAAA 21
```

```
RESULT 5
US-09-837-604A-1465/C
; Sequence 1465, Application US/09837604A
; GENERAL INFORMATION:
; APPLICANT: Byrum, Joseph R.
; APPLICANT: La Rosa, Thomas J.
; APPLICANT: Shukla, Hridayabhiranjan
; APPLICANT: De la Pena, Robert C.
; APPLICANT: Bougri, Olegs
; TITLE OF INVENTION: Nucleic Acid Molecules And Other Molecules Associated With
; FILE REFERENCE: 38-21(51892)B
; CURRENT APPLICATION NUMBER: US/09/837,604A
```

```

; CURRENT FILING DATE: 2001-04-18
; PRIOR APPLICATION NUMBER: US 60/197,872
; PRIOR FILING DATE: 2000-04-19
; NUMBER OF SEQ ID NOS: 81288
; SEQ ID NO 1465
; LENGTH: 592
; TYPE: DNA
; ORGANISM: Oryza sativa nippobare
; FEATURE:
; OTHER INFORMATION: Clone ID: LIB3431-048-P1-N1-G3
US-09-837-604A-1465

```

```

Query Match          5.08; Score 54.8; DB 7; Length 592;
Best Local Similarity 50.38; Pred. No. 0.45;
Matches 163; Conservative 0; Mismatches 157; Indels 4; Gaps 1;

```

```

QY 46 TTCCAAATTTTGTGTAGCAAACTCCAAATTTTGTGTGTAGCAACAGAAA 105
   || ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 460 TTTTAAAAAAATTTTATTTTAAAAAAAGTTTGTGTGTATATATAGTTT 401
   || ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 106 TTAACAGATGAAACATATTTGTGTGGAATGGAATGATATATACATTAACCAATTT 165
   || || ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 400 GTGACGTTTAAATATATTTTGTGAAAAAAGAGATTAATATTTT 341
   || || ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 166 TAAAAATTTATTAACCTATACGCGT---CAAGTATGTATCTAGTAGGTATTT 221
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 340 TAAAAAATTTTGTACAAAAATATTTTCAAAAAATTTTGTGATTTT 281
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 222 AATAATGATGTCGCTTCAATGGAATGGCAACATGAAGGATTAATATTAAC 281
   || || ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 280 AAAAAAAGCGCAAAATTTTGTAGTTAAAAAACCATATTAATATTTT 221
   || ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 282 TTTAAAAATATTAATTTGAGTAAATGTGTCTGACTATTTGAGGCAAAAAAG 341
   || ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 220 TTTCAAAAAAATTTTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 161
   || ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 342 ACAATGCCAAATCTACGGGTTT 365
   || ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 160 AATTTTAAAAAATTTTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 137
   || ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

```

RESULT 6

```

US-60-466-412-84116
; Sequence 84116; Application US/60466412
; GENERAL INFORMATION:
; APPLICANT: CARGILL, Michele
; APPLICANT: IAKOUBOVA, Olga
; TITLE OF INVENTION: GENETIC POLYMORPHISMS ASSOCIATED WITH
; TITLE OF INVENTION: MYOCARDIAL INFARCTION, METHODS OF DETECTION AND USES THEREOF
; FILE REFERENCE: CL001466
; CURRENT APPLICATION NUMBER: US/60/466,412
; CURRENT FILING DATE: 2003-04-30
; NUMBER OF SEQ ID NOS: 429241
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 84116
; LENGTH: 23939
; TYPE: DNA
; ORGANISM: Homo sapiens
US-60-466-412-84116

```

```

Query Match          4.98; Score 54.4; DB 10; Length 23939;
Best Local Similarity 50.08; Pred. No. 0.45;
Matches 163; Conservative 0; Mismatches 161; Indels 2; Gaps 1;

```

```

QY 1 AATCTCAATTTTATATATTTAGTCAATGATGATGATGATGATGATGATGATGAT 60
   || || ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 23107 AATATTTATATATTTATATATTTATATATATATATATATATATATATATATAT 23166
   || || ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 61 TGTAAACAGAACTTCCAAATTTTGTGTGTGGAACAGAAATTAACAGATAGAAAC 120
   || || ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 23167 TATATATATATATATATATATATATATATATATATATATATATATATATATAT 23226
   || || ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 121 TATTTGTGTGGAATGAGTAGTATATATATATATATATATATATATATATATAT 180
   || || ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

```

```

DB 23227 TTAATAGATTAATGATTTATATATATATATATATATATATATATATATATAT 23286
   || || ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 181 GCCATAGCGCGCTCAAGATGATGATGATGATGATGATGATGATGATGATGATGAT 238
   || || ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 23287 TTAATAGATTAATGATTTATATATATATATATATATATATATATATATATAT 23346
   || || ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 239 TTTCAAGATTTGGACACACAGAAACGAAATTAATATATATATATATATATATATAT 298
   || || ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 23347 ATAATATTAATGATTAATATATATATATATATATATATATATATATATATAT 23406
   || || ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 299 TTTGATTAATGCTTTTGTGACTAT 324
   || || ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 23407 TTAATATTAATATATATATATATATATATATATATATATATATATATATAT 23432
   || || ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

```

RESULT 7

```

US-09-912-293-46414/C
; Sequence 46414; Application US/09912293
; GENERAL INFORMATION:
; APPLICANT: Rosen, et. al.
; TITLE OF INVENTION: Human Genes, Sequences, and Expression Products 100
; FILE REFERENCE: PO-100
; CURRENT APPLICATION NUMBER: US/09/912,293
; CURRENT FILING DATE: 2001-07-26
; PRIOR APPLICATION NUMBER: 08/103,744
; PRIOR FILING DATE: 1993-08-09
; PRIOR APPLICATION NUMBER: 09/249,651
; PRIOR FILING DATE: 1999-02-12
; PRIOR APPLICATION NUMBER: 08/104,507
; PRIOR FILING DATE: 1993-08-09
; PRIOR APPLICATION NUMBER: 08/196,363
; PRIOR FILING DATE: 1994-02-15
; PRIOR APPLICATION NUMBER: 09/859,490
; PRIOR FILING DATE: 2001-05-18
; PRIOR APPLICATION NUMBER: 08/196,362
; PRIOR FILING DATE: 1994-02-15
; PRIOR APPLICATION NUMBER: 08/221,623
; PRIOR FILING DATE: 1994-03-31
; PRIOR APPLICATION NUMBER: 08/220,691
; PRIOR FILING DATE: 1994-03-31
; PRIOR APPLICATION NUMBER: 09/741,830
; PRIOR FILING DATE: 2000-12-22
; PRIOR APPLICATION NUMBER: 09/813,155
; PRIOR FILING DATE: 2001-03-21
; Remaining Prior Application data removed - See File Wrapper or PALM.
; NUMBER OF SEQ ID NOS: 244538
; SEQ ID NO 46414
; LENGTH: 427
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (3)..(3)
; OTHER INFORMATION: n is equal to a,t,g, or c
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (16)..(16)
; OTHER INFORMATION: n is equal to a,t,g, or c
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (27)..(27)
; OTHER INFORMATION: n is equal to a,t,g, or c
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (36)..(36)
; OTHER INFORMATION: n is equal to a,t,g, or c
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (77)..(77)
; OTHER INFORMATION: n is equal to a,t,g, or c
; FEATURE:
; NAME/KEY: misc_feature

```



```

; PRIOR APPLICATION NUMBER: PCT/EP01/07537
; PRIOR FILING DATE: 2001-07-02
; PRIOR APPLICATION NUMBER: DE 10032529.7
; PRIOR FILING DATE: 2000-06-30
; PRIOR APPLICATION NUMBER: DE 10043826.1
; PRIOR FILING DATE: 2000-09-01
; NUMBER OF SEQ ID NOS: 2424
; SEQ ID NO 776
; LENGTH: 6113
; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: chemically treated genomic DNA (Homo sapiens)
US-10-311-455-776

```

```

Query Match
Best Local Similarity 51.9%; Score 54.2; DB 9; Length 6113;
Matches 122; Conservative 0; Mismatches 113; Indels 0; Gaps 0;

```

```

QY 6 ACAATTTATTTATTTAGTCATATACATGACATGAGAAAGTCCAAAAAATTTTGTGA 65
DB 5823 ACACAAATCTAAAAATATAAAAAACAATACTTAAATCTCATATAAAAAATTTATCA 5764
QY 66 ACAGAACTCCAAATTTTATTTATGGAACAAGAAATACAGATAGAAAGTATT 125
DB 5763 AAAAAAATCTTAAATTTATTTATTTAACTTAAAAAATTTTATTTATTTAT 5704
QY 126 TGTGTGCAATGAGTAAATATATATATAGCAAAATTTTAAAAATTTATATAGCCTA 185
DB 5703 TCTCTTAAAAAATATCATATAATATAAATATATATTTTATTTATACATCTA 5644
QY 186 TACGCGCTCAAGATGATTTATCTAGTAGTGTAATTAATATGATGCGGAT 240
DB 5643 CGTATCTATATAATCTTTTATTTTATTTTAAAAATATAATCTCGCCCTATT 5589

```

```

RESULT 10
US-60-466-412-86174
; Sequence 86174, Application US/60466412
; GENERAL INFORMATION:
; APPLICANT: CARGILL, Michele
; APPLICANT: IAKOUBOVA, Olga
; TITLE OF INVENTION: GENETIC POLYMORPHISMS ASSOCIATED WITH
; TITLE OF INVENTION: MYOCARDIAL INFARCTION, METHODS OF DETECTION AND USES THEREOF
; FILE REFERENCE: CI001466
; CURRENT APPLICATION NUMBER: US/60/466,412
; CURRENT FILING DATE: 2003-04-30
; NUMBER OF SEQ ID NOS: 429241
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 86174
; LENGTH: 148843
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (1)..(148843)
; OTHER INFORMATION: n = A,T,C or G
US-60-466-412-86174

```

```

Query Match
Best Local Similarity 50.2%; Score 54.2; DB 10; Length 148843;
Matches 157; Conservative 1; Mismatches 154; Indels 1; Gaps 1;

```

```

QY 6 ACAATTTATTTATTTAGTCATATACATGACATGAGAAAGTCCAAAAAATTTTGTGA 65
DB 102501 ATATATATATATATATATATATATATATATATATATATATATATATATATATAT 102560
QY 66 ACAGAACTCCAAATTTTATTTATGGA-ACAAGAAATACAGATAGAAAGTATT 124
DB 102561 AATATATATATATATATATATATATATATATATATATATATATATATATATATAT 102620
QY 125 TGTGTGCAATGAGTAAATATATATATAGCAAAATTTTAAAAATTTATATAGCCT 184

```

```

DB 102621 TTAATATATATATATATATATATATATATATATATATATATATATATATAT 102680
QY 185 ATACGCGTCAAGATGATTTATCTAGTGTATATATATATATATATATATATATATAT 244
DB 102681 ACATCTTATATAGTATATATATATATATATATATATATATATATATATATATAT 102740
QY 245 ATGGGACACAGTAAAGGAAATTTAAATATATATATATATATATATATATATATAT 304
DB 102741 TAAATATATATATATATATATATATATATATATATATATATATATATATAT 102800
QY 305 TAAATGCTTTTC 317
DB 102801 AAGTTGCTTTTC 102813

```

```

RESULT 11
US-09-806-708B-22/C
; Sequence 22, Application US/09806708B
; GENERAL INFORMATION:
; APPLICANT: The University of British Columbia
; TITLE OF INVENTION: Regulation of Embryonic Transcription in Plants
; FILE REFERENCE: 4810-58741
; CURRENT APPLICATION NUMBER: US/09/806,708B
; PRIOR FILING DATE: 2001-04-03
; PRIOR APPLICATION NUMBER: US 60/147,133
; PRIOR FILING DATE: 1999-08-04
; NUMBER OF SEQ ID NOS: 23
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 22
; LENGTH: 1141
; TYPE: DNA
; ORGANISM: Artificial sequence
; FEATURE:
; NAME/KEY: promoter
; LOCATION: (1)..(1141)
; OTHER INFORMATION: consensus sequence of A.t., L.a., and B.n. FAEI promoters
US-09-806-708B-22

```

```

Query Match
Best Local Similarity 11.3%; Score 53.8; DB 6; Length 1141;
Matches 110; Conservative 343; Mismatches 516; Indels 5; Gaps 2;

```

```

QY 74 TTCGAATTTTATTTTATGGAACAAGAAATACAGATGAGAAAGTATTTTGTGCG 133
DB 1125 KTYKANNNNNNNNNGMDNRMDATKWSATGATMTNHAARGATMCWYWTGTRRMC 1066
QY 134 AATGAGTACTAATATATATATAGCAAAATTTTAAAAATTTATATATATATATATATAT 193
DB 1065 MRYTAMRWYTRSNANNSCATKBMMWTKYATKRTMYAMICAMRNMMCATNGYAK 1006
QY 194 CAAAGTATGTTATCTAGTGTGATTAATTAATGATGCGATTCGATTCAGATTTGGACA 253
DB 1005 SCATNNMMWYATTTAAAYAAAKWMAAGNNNRBMGAAGNNKWCMAAATGMBMADTAG 946
QY 254 ACAATGAAACGCAATTTAAATATATATATATATATATATATATATATATATATATAT 313
DB 945 KMCNNNNNNNTTDDVRMAMKAKNNNNNNNAWYACYNRAATNNKMATHHMKWTHGASHKR 886
QY 314 TTTCGACTATGAGGCGCAAAAAAAGACATGCGCAAAAGCTACGCGTTTGAAGTCC 373
DB 885 TRHRTCTCRKRYNNNNNNNATVTYVYHHAARMMAMATRTNNNNNNNNNACRTKRTW 826
QY 374 AGTTCGGTATATATATATATATATATATATATATATATATATATATATATATATAT 433
DB 825 ABMKHSMNNNNNNNNNNNNNT---WCHYTANABDLYRANNNNAAARMAATCANNYHA 769
QY 434 ACATTTCTAGTCTACCCCTACTCGTAGCCACCCCTTTCCATATCTAGAGGTA 493
DB 768 AVTTTHDWCYKTMNTWYMDMTTBTTRNNMTTSTNNMTNNNNNNNMACTNNNNNNMM 709
QY 494 TTTTGAATCCCAATTTAAACGATGAGACCGTACCGGACTCTCGGATTCGCGG 553
DB 708 KAYTAHATNNMGCMNNNTDARTNTVYMRMRMTNTKRTYSTRHRYHGTATNNNNN 649

```


GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: June 10, 2003, 22:04:49 ; Search time 1669.1 Seconds

(without alignments)
10683.154 Million cell updates/sec

Title: US-09-502-426b-1_COPY_2102_3202

Perfect score: 1101

Sequence: 1 aatcaacaattataat.....gagagagagaactagctcc 1101

Scoring table: IDENTITY NUC

Gapop 10.0, Gapext 1.0

Searched: 16154066 seqs, 8097743376 residues

Total number of hits satisfying chosen parameters: 32308132

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Database : Listing first 45 summaries

EST:*
1: em_estba:*
2: em_esthum:*
3: em_estin:*
4: em_estmu:*
5: em_estov:*
6: em_estpl:*
7: em_estro:*
8: em_esti:*
9: gb_est1:*
10: gb_est2:*
11: gb_est3:*
12: gb_est4:*
13: gb_est5:*
14: em_estfun:*
15: em_estom:*
16: gb_gss:*
17: em_gss_hum:*
18: em_gss_inv:*
19: em_gss_pln:*
20: em_gss_vit:*
21: em_gss_fun:*
22: em_gss_mam:*
23: em_gss_mus:*
24: em_gss_other:*
25: em_gss_pro:*
26: em_gss_rod:*
27: em_gss_rod:

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
c 1	148	13.4	148	17	BH811044
2	145.6	13.2	473	17	AL768954 Arabidops
3	69.6	6.3	1101	17	CNS0039G
c 4	69.4	6.3	1201	17	CNS0167M
5	66.8	6.1	928	17	CNS00DKY
c 6	66.4	6.0	1201	17	CNS0165X

7	65.8	6.0	1101	17	CNS0021J	AL061936 Drosophila
8	65.2	5.9	1101	17	CNS00172	AL078714 Drosophila
c 9	64.6	5.9	1101	17	CNS000EV	AL069706 Drosophila
10	63.6	5.8	1146	17	CNS021G2	AL176843 Tetradon
c 11	63.4	5.8	1101	17	CNS0021J	AL061936 Drosophila
c 12	63.2	5.7	1101	17	CNS00B01	AL057419 Drosophila
c 13	62.8	5.7	1101	17	CNS00587	AL057153 Drosophila
c 14	62.6	5.7	1101	17	CNS000EV	AL057153 Drosophila
c 15	62.4	5.7	855	17	CNS04P5D	AL300874 Tetradon
c 16	62	5.6	1029	17	CNS01ZGM	AL174271 Tetradon
c 17	61.8	5.6	1101	17	CNS003BD	AL064091 Drosophila
c 18	61.6	5.6	609	17	CNS025K2	AL182171 Tetradon
c 19	61.6	5.6	1101	17	CNS0161I	AL106896 Drosophila
c 20	61.6	5.6	1190	17	CNS020M7	AL206908 Tetradon
c 21	61.6	5.6	1200	17	CNS016CO	AL106578 Drosophila
c 22	61.4	5.6	1092	17	CNS020K7	AL142826 Tetradon
c 23	61	5.5	910	17	CNS01G8P	AL142826 Tetradon
c 24	61	5.5	1085	17	AG075009	AG075009 Pan trogl
c 25	60.8	5.5	1101	17	CNS00B01	AL057419 Drosophila
c 26	60.8	5.5	1101	17	CNS001FB	AL060732 Drosophila
c 27	60.6	5.5	878	17	CNS0187R	AL108993 Drosophila
c 28	60.6	5.5	1101	17	CNS0039G	AL063921 Drosophila
c 29	60.2	5.5	1201	17	CNS016C2	AL106556 Drosophila
c 30	60	5.4	945	17	CNS04D0K	AL285149 Tetradon
c 31	59.8	5.4	1225	17	CNS0161D	AL106171 Drosophila
c 32	59.8	5.4	964	17	CNS07E8R	AL441457 T7 end of
c 33	59.8	5.4	966	17	CNS0052C	AL061991 Drosophila
c 34	59.6	5.4	1092	17	CNS020K7	AL175696 Tetradon
c 35	59.6	5.4	1368	10	BE420618	BE420618 HMM000.G0
c 36	59.2	5.4	524	17	CNS01090	AL167541 Tetradon
c 37	59.2	5.4	843	17	CNS00CS1	AL055666 Drosophila
c 38	59	5.4	987	17	CNS004PO	AL054456 Drosophila
c 39	59	5.4	994	17	CNS04NOJ	AL298902 Tetradon
c 40	58.8	5.3	1101	17	CNS008X3	AL052544 Drosophila
c 41	58.6	5.3	1001	17	CNS0155H	AL105023 Drosophila
c 42	58.6	5.3	1101	17	CNS00EO7	AL069440 Drosophila
c 43	58.4	5.3	676	17	CNS003XU	AL065304 Drosophila
c 44	58.4	5.3	843	17	CNS00CS1	AL059666 Drosophila
c 45	58.4	5.3	996	17	CNS00F0H	AL071063 Drosophila

ALIGNMENTS

RESULT 1
LOCUS BH811044 148 bp DNA linear GSS 02-MAY-2002
DEFINITION SALK_057128 Arabidopsis thaliana TDNA insertion lines Arabidopsis
thaliana genomic clone SALK_057128, DNA sequence.
BH811044
ACCSSION BH811044.1 GI:20388862
VERSION
KEYWORDS
SOURCE
ORGANISM
GSS.

thale cress,
Arabidopsis thaliana

Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
Rosidae; eurosids II; Brassicales; Brassicaceae; Arabidopsis.
1 (bases 1 to 148)

Alonso,J.M., Leisse,T.J., Barajas,P., Chen,H., Cheuk,R., Gadrinab,
,C., Jeske,A., Karnes,M., Kim,C.J., Parker,H., Prednis,L., Shinn,P.,
, Zimmerman,J., and Ecker,J.R.
A Sequence-Indexed Library of Insertion Mutations in the
Arabidopsis Genome

Unpublished (2001)

JOURNAL

COMMENT

Contact: Joseph R. Ecker
Salk Institute Genomic Analysis Laboratory (Signal)
The Salk Institute for Biological Studies
10010 N. Torrey Pines Road, La Jolla, CA 92037, USA
Tel: 858 453 4100 x1752
Fax: 858 558 6379
Email: eckeresalk.edu

This is single pass sequence recovered from the left border of
TDNA.

Class: TDNA tagged.
Location/Qualifiers
1. 148
/organism="Arabidopsis thaliana"
/strain="Columbia 0"
/db.xref="taxon:3702"
/clone="SALK_057128"
/note="PCR was performed on Arabidopsis thaliana lines each of which contains one or more TDNA insertion elements. The resultant fragment for each line was directly sequenced to determine the genomic sequence at the site of insertion. Details of the protocols used can be found at http://signal.salk.edu/tdna_protocols.html"

BASE COUNT 46 a 28 c 14 g 60 t
ORIGIN

Query Match 13.4%; Score 148; DB 17; Length 148;
Best Local Similarity 100.0%; Pred. No. 6.6e-18;
Matches 148; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 93 ATGCACAGAAATACAGATAGAACTATTTGTTGTGATGAGTAAATATAC 152
DB 148 ATGGAACAGAAATACAGATAGAACTATTTGTTGTGATGAGTAAATATAC 89
QY 153 ATTAGCAATTTTAAAAATTTATATAGCCTATAGCGCTCAAGTATCTAGTA 212
DB 88 ATTAGCAATTTTAAAAATTTATATAGCCTATAGCGCTCAAGTATCTAGTA 29
QY 213 GGTGTATTAATATGATGATGCGGATT 240
DB 28 GGTGTATTAATATGATGATGCGGATT 1

RESULT 2 473 bp DNA linear GSS 18-JUN-2002
AL768954 Arabidopsis thaliana T-DNA flanking sequence GK-082A08-011867,
LOCUS genomic survey sequence.
ACCESSION AL768954
VERSION AL768954.1 GI:21522073
KEYWORDS GSS.
SOURCE thale cress.
ORGANISM Arabidopsis thaliana
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae; eurosids II; Brassicales; Brassicaceae; Arabidopsis.

REFERENCE 1
Sritihov, N., Li, Y., Rosso, M., Viehoveer, P., Dekker, K., Saedler, H. and Weisshaar, B.
TITLE A pipeline for automated high-throughput generation of FSTs (flanking sequence tags) from Arabidopsis thaliana T-DNA transformed lines
JOURNAL Unpublished
AUTHORS 2
Rosso, M., Sritihov, N., Li, Y., Reiss, B., Dekker, K. and Weisshaar, B.
TITLE A new Arabidopsis thaliana T-DNA mutagenised population (GABI-Kat) for flanking sequence tag based reverse genetics
JOURNAL Unpublished
AUTHORS 3 (bases 1 to 473)
Li, Y., Rosso, M., Sritihov, N. and Weisshaar, B.
COMMENT Direct Submission
Submitted (17-JUN-2002) Weisshaar B., Max-Planck-Institut fuer Zuechtungsforchung, Carl-von-Linne-Weg 10, Koeln, 50829, Germany
This sequence is recovered from the left border of the T-DNA. It indicates an insertion within the locus defined by clone T3A5. The sequences are generated at the MPI for Plant Breeding Research in the context of the GABI-Kat project. GABI-Kat is part of the German Plant Genomics program designated 'GABI'. Information on line availability can be found at:
<http://www.mpi-z-koeln.mpg.de/GABI-Kat/>.
location/Qualifiers
1. 473

/organism="Arabidopsis thaliana"
/strain="Columbia 0"
/db.xref="taxon:3702"
/clone="GK-082A08-011867"
/note="PCR was performed on DNA from Arabidopsis thaliana plants (T1) which were transformed with the T-DNA from vector pAC161. The lines contain one or more T-DNA insertions. The DNA fragment(s) resulting from the PCR were directly sequenced to determine the genomic sequence flanking the insertion. Sequences displaying significant similarity to the A. thaliana nuclear genome sequence were processed for submission. T-DNA derived sequences were removed"

BASE COUNT 168 a 58 c 70 g 146 t 31 others
ORIGIN

Query Match 13.2%; Score 145.6; DB 17; Length 473;
Best Local Similarity 87.0%; Pred. No. 1.3e-17;
Matches 160; Conservative 0; Mismatches 24; Indels 0; Gaps 0;

QY 1 AATCTACAAATTTATATATTTAGTCATTAACATGCAATGCAAAAGTCCAAAATTT 60
DB 290 AATCTACAAATTTATATATTTAGTCATTAACATGCAATGCAAAAGTCCAAAATTT 349
QY 61 TGTTAACAGAAATTCCTCAATTTTTTTTTTATGCAACAGAAATTAACAGATGAAAC 120
DB 350 TGTTAACAGAAATTCCTCAATTTTTTTTTTGGGGCCAAAATTAACCGATGAAAC 409
QY 121 TATTTGTTGGAATGGAATGATTAATATCATTAAGCAATTTAAATTTATATA 180
DB 410 TTTTGGGGGGGAGGAGGAGTATTAATATCTTAACCAATTTTAAAAATTTTAA 469
QY 181 GCTT 184
DB 470 CCTT 473

RESULT 3 1101 bp DNA linear GSS 03-JUN-1999
CNS0039G Drosophila melanogaster genome survey sequence TET3 end of BAC #
LOCUS BACR08K10 of RPCR-98 library from Drosophila melanogaster (fruit fly), genomic survey sequence.
ACCESSION AL063921
VERSION AL063921.1 GI:4941778
KEYWORDS GSS.
SOURCE Drosophila melanogaster.
ORGANISM Drosophila melanogaster.
Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.

REFERENCE 1 (bases 1 to 1101)
Genoscope.
TITLE Direct Submission
JOURNAL Submitted (02-JUN-1999) Genoscope - Centre National de Sequencage : BP 191 91006 Evry cedex - FRANCE (E-mail : seqref@genoscope.cns.fr - Web : www.genoscope.cns.fr)
COMMENT determination of this BAC-end sequence was carried out as part of a collaboration with the Berkeley Drosophila Genome project (BDGP). The BDGP is constructing a physical map of the Drosophila melanogaster genome using these BACs. For further information please see <http://www.fruitfly.org> The BDGP Drosophila melanogaster BAC library was prepared by Kazuo Osoegawa and Aaron Margomer in Pletier de Jong's laboratory in the Department of Cancer Genetics at the Roswell Park Cancer Institute in Buffalo, NY. The library is named RPCR-98 and was constructed by partial EcoRI digestion of Drosophila DNA provided by the BDGP from the isogenic strain Y2; cn bw sp, the same strain used for the BDGP's p1 and EST libraries. A more detailed description of the library and how to order individual BAC clones, the entire library, or filters for hybridization from the BACPAC Resource Center can be found at http://bacpac.med.buffalo.edu/drosophila_bac.htm.

Location/Qualifiers
1. .1101

BASE COUNT	201 a	64 c	131 g	202 t	503 others
ORIGIN					

Query Match	6.38;	Score 69.6;	DB 17;	Length 1101;
Best Local Similarity	19.28;	Pred. No. 0.0016;		
Matches	97;	Conservative 218;	Mismatches 107;	

[illegible]

1. .1201

<http://www.edgp.ebi.ac.uk> - This Drosophila melanogaster BAC library (Dros_BAC) was made by Alain Billand at CEPH (Centre d'Etude du Polymorphisme Humain) with funding provided by a MRC project grant. The DNA was prepared from embryos by Alain Bucheton and Genevieve Payan. It has been constructed in the vector pBlaoBAC11.

BASE COUNT	323 a	87 c	79 g	551 t	161 others
ORIGIN					

Query Match	6.3%;	Score 69.4;	DB 17;	Length 1201;
Best Local Similarity	37.2%;	Pred. No. 0.0017;		
Matches 200;	Conservative 57;	Mismatches 280;	Indels 0;	Cross 0

```

OY 557 ATTATCAAAAATTTATTAGCAGCAATGGGTTTATTTAATTTAAAAAAGCTCAACTGATGCA 616
Db 1147 AANATATAAAAAAMAAAATATAMWATATTAANAAMAAAAMWMAATATTTAAATWMAATATANA 1088
OY 617 GATPAAAATTTCAATAACACTTTTACGATGGATTCGACATGATATCTATAGCTTTT 676
Db 1087 AAAAAAAMAAAAMWMTTTHHANATTTTWTWTTWNTATWMAATWTTT 1028
OY 677 TTCTCAACAGGATGAGTGAAGTTATATAGTATTTAGCCAGACAAATGATTTATAGATA 736
Db 1027 TTTATATAMAAAAMAAAATTTTAAAAATTAATTAATTAATTAATTTTAAAAAT 968
OY 737 TAATCCATTAATCCAGATATTTATGATATAAATAGCTGTAAACATTTGACATCGGAG 796
Db 967 TTTWATWMTTTTAAAAAAMAAAATAWMAATTTTTTTTATWATAAAMWMTT 908
OY 797 CTTTTCGAACCTTTGTTTTTAATTTAAGAGTTAATTAATAAAGATTTAAAAAGGACA 856
Db 907 TTTTWTAAAAAAMAAAAMTTAAAAATTTAAAAATTTAAAAAAMAAAAMAAAAA 848
OY 857 TAACGAGCAACAAAAGTAATGACACGAGAAAACAAAAGCATGAAGCTCATGTGGTAG 916
Db 847 AAAAAAAMAAAAMTTWATATTAAMATTTTAAAAAAMAAAAMAAAATTTAATTTTAAAA 788
OY 917 TTTAAGCTTAATAGAGAGATTTTATTAATTTTAAGACATGATATACATTTATTTTC 976
Db 787 ATAAAAAAMAAAAMWMAAAAAATWMTAATATATWTTAAATCATTAACAAAAAAMAA 728
OY 977 TGACTCTTTAAAAACCCCTTTTCAACAGAGAGCTCCCTTTTCAGTAGAAGCTCGATT 1036
Db 727 TTTWATTAATAAAAAAAMAAAAMAAAAMAAAAMAAAAMAAAAMAAAAMAAAAMAAAAM 668
OY 1037 CCAATCTTAAAGACAAAGCATTTAGAAAGAGAGATGAGTGAAGAGAGAGAA 1093
Db 667 MCAAAAAAMAAAAMAAAAMCAAMMTTAAAGCTTAAAAAACAAMAAAAMAAAAMAAA 611

```

melanogaster:

Accession: AL106396
Version: AL106396.1
Keywords: GSS
GI:5621701

REFERENCE
1 (bases 1 to 1201)
Ephydroidae; Drosophilidae; Drosophila.
Muscomorpha;
Diptera; Brachycera; Muscomorpha;
Ephydroidae; Drosophilidae; Drosophila.

06 EVERY CE

Genoscope.
AUTHORS
TITLE
Direct Submission
Submitted (23-JUL-1999) Genoscope - Catherine

on with the European Commission as part of a

CC 06 / MC AAAAAMAAA VMAAAACMAMMTTAAAGCTAAAAAACCAAAAAMAAAATA 611

GT:4948170

ACCESSION
AL071865
BACR27A24 of RPCI-98 library from *Drosophila melanogaster* (fruit fly), genomic survey sequence.

melanogaste

```

VERSION      AL071865.1  GI:4948170
KEYWORDS     GSS.
SOURCE       HSC030411-01

```

Endopterygota; Diptera; Brachycera; Muscomorpha;

- Web : www.genoscope.cns.fr

- Web : www.genoscope.cns.fr
Determination of this BAC-end sequence was carried out as part of a collaboration with the Berkeley Drosophila Genome Project (BDGP).
The BDGP is constructing a physical map of the Drosophila melanogaster genome using these BACs. For further information please see <http://www.fruitfly.org>
The BDGP Drosophila melanogaster BAC library was prepared by Kazutoyo Osoegawa and Aaron Mamonosier in Plater de Jong's laboratory in the Department of Cancer Genetics at the Roswell Park Cancer Institute in Buffalo, NY. The library is named RPCI-98 and was constructed by partial EcoRI digestion of Drosophila DNA provided by the BDGP from the isogenic strain y2; cn bw sp, the same strain used for the BDGP's P1 and EST libraries. A more detailed description of the library and how to order individual BAC clones, the entire library, or filters for hybridization from the BACPAC Resource Center can be found at http://bacpac.med.buffalo.edu/drosophila_bac.htm.

1. .1101

/organism="Drosophila melanogaster"

631 a	7 c	28 g	289 t	146 others
-------	-----	------	-------	------------

Query Match	6.0%;	Score 65.8;	DB 17;	Length 1101;
Best Local Similarity	43.5%;	Pred. No. 0.0082;		
Matches 136;	Conservative 30;	Mismatches 147;	Indels 0;	Gaps 0;

[illegible]

CNS00LFR2 1101 bp DNA linear GSS 14-JUN-1999
Drosophila melanogaster genome survey sequence TERT3 end of BAC:
BACR48B19 of RPCI-98 library from Drosophila melanogaster (fruit
fly), genomic survey sequence.
AL078714
AL078714.1 GI:5102004
GSS:
Drosophila melanogaster.
Drosophila melanogaster.

REFERENCE
AUTHORS
TITLE
JOURNAL
Submitted (11-JUN-1999) Genoscope - Centre National de Séquençage :
1 (bases 1 to 1101)
Genoscope.
Direct Submission
Epigenetica, Chromophilaecae, Chromophila.

BP 191 91006 EVRY cedex - FRANCE (E-mail : segref@genoscope.cns.fr)
- Web : www.genoscope.cns.fr)
Determination of this BAC-end sequence was carried out as part of a

BP 191 91000 EVRY cedex - FRANCE (E-mail : sequire@genoscope.cns.fr
Web : www.genoscope.cns.fr)
Determination of this BAC-end sequence was carried out as part of a
collaboration with the Berkeley Drosophila Genome Project (BDGP).
The BDGP is constructing a physical map of the Drosophila
melanogaster genome using these BACs. For further information
please see <http://www.fruitfly.org> The BDGP Drosophila
melanogaster BAC library was prepared by Kazutoyo Osoegawa and
Ganon Mammets in Pleier de Jong's Laboratory in the Department of
Cancer Genetics at the Roswell Park Cancer Institute in Buffalo,
NY. The library is named RPc1-98 and was constructed by partial
EcoRI digestion of Drosophila DNA provided by the BDGP from the
isogenic strain y; cn bw sp, the same strain used for the BDGP's
p1 and EST libraries. A more detailed description of the library
and how to order individual BAC clones, the entire library, or
filters for hybridization from the BACPAC Resource center can be
found at http://bacpac.med.buffalo.edu/bacpac/drosophila_bac.htm.

1. .1101

```
/organism="Drosophila melanogaster"
/db_xref="taxon:72327"
```

469 a	6 c	69 g	151 t	406 others
-------	-----	------	-------	------------

Query Match	5.9%	Score	65.2;	DB	17;	Length	1101;
Best Local Similarity	32.0%;	Pred. No.	0.011;				
Matches 133; Conservative	90;	Mismatches	193;	Indels	0;	Gaps	0;

QY	564	AAAAATATATGACGCAATGGGTTTATTAATTTTAAAAAGCACAACCTGGATCAAT	623
Db	418	AA	477
QY	624	TTTCATTAACACTTTTACGATTCGATTCGATTCGATTCGATTCGATTCGATTCGATTC	683
Db	478	AAAAAAAAAATTTTATTTTATTTTATTTTATTTTATTTTATTTTATTTTATTTTATTT	537
QY	684	CACGGTGGATGAAAGTTATAGTACTATTAGCCGAGACAATTGATTTTACGATTAATCCAT	743
Db	538	TAAATATTTTATTTTATTTTATTTTATTTTATTTTATTTTATTTTATTTTATTTTATTT	597
QY	744	TAAATCATGATATTATATGATATATAATATGCGTTTAATATATTTCAGCATCCAGCTTCTG	803
Db	598	TAAATTAATTTTATTTTATTTTATTTTATTTTATTTTATTTTATTTTATTTTATTTTATTT	657
QY	804	CAACTTTTGGTTTAAATTTTAAGAGTTAATAAATAAAGATTTAAAGGAGCATTAACGAG	863
Db	658	TATCTATTTTATTTTATTTAATTAATAAATAAATTTTAAATAAATTTTATTTTATTTTATTT	717
QY	864	GCAACAAAAGTAAATGAAACGAGGAAACAAAGCATCAAGCTCATGTTAGTTAAGC	923
Db	718	KMGAGAKATTTTCKKDKKAAAMAADKMKRKKGGKGGKGGKGGKGGKGGKGGKGGKGG	777
QY	924	TTAATAAGAAAGATTTTAAATTTTAAATGACAGTGAATACAAATTAATTTTCTGA	979
	778	GGKKKKGGDDADKXTKKKKKKAAATTTTCKKKKKKKKKKKKKKKAAADTCTTKKDA	833

CNS00EVL 1101 bp DNA linear GSS 04-JUN-1999
Drosophila melanogaster genome survey sequence T7 end of BAC:
BACR9823 of RPCI-98 library from Drosophila melanogaster (fruit
[iv]), genomic survey sequence.

KEYWORDS
SOURCE
ORGANISM
GSS.
Drosophila melanogaster.
Drosophila melanogaster.
Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;

Ephydroidea; Drosophilidae; Drosophila.
1 (bases 1 to 1101)

TITLE	JOURNAL	COMMENT
Direct Submission	Submitted (02-JUN-1999) Genoscope - Centre National de Sequencage : BP 191 91006 EVRY cedex - FRANCE (E-mail : sequefegenoscope.cns.fr - Web : www.genoscope.cns.fr)	Determination of this BAC-end sequence was carried out as part of a

FEATURES	Location/Qualifiers
source	1. .1101

BASE COUNT	419 a	91 c	60 g	299 t	232 others
ORIGIN					

Query Match	5.9%	Score 64.6	DB 17	Length 1101
Best Local Similarity	34.3%	Pred. No. 0	014	
Matches 139	Conservative 71	Mismatches 195	Indels 0	Gaps 0

QY	586	TTTATATATTTAAAACTCACACCTGTCAGATAAAATTCATTAACACCTTTCAGATG	645
Db	915	TTTATATTTTATATATTTAAAAAAMAAAMTATTTTATTTTATATATTAATATTTAA	856
QY	646	GATTCGACATCTATCTAATGACTTTTTTTTTTCTACACGCGTGATGAAGTATAGT	705
Db	855	TTTTTTTWTATATTTTATATATATATTTTATATTTTATATATATATATATATTTTWT	796
QY	706	ACTATTAGCAGACAAATGATATATAGATATATCCATTAATCCATGATATTTATGAT	765
Db	795	ATTWTATATTTTAAATTTTAAATTTTAAATTTTAAATTTTAAATTTTAAATTTT	736
QY	766	AAATAGCTGTAAACATTTTCAGACATGACGACCTTTCGACACTTTGTTTATATTTAA	825
Db	735	ATTWTTTTAAATATATATATATATTTTATATTTTATATTTTATATTTTAAATTTT	676
QY	826	AGTTTATATATATAAAGTATTTAAAAAGAGCATTAACGAGCAACAAAGTAATGACACGG	885
Db	675	TAAATTTAAATATATATTTTAAATTAATTAATTAATTAATTAATTAATTAATTAATTA	616
QY	886	AGAAACAAAAGCCATGAGCTCATGTTGGTTAGTTAGCTTAATTAAGAAAGATTTATTA	945
Db	615	TTTTTTTATATATATATATTTTATATATTTTATATATATATATATATATTTTAAATTT	556
QY	946	TTTTATGACGATGATTAACATATATATTTCTGCTCTTTAA	990
Db	555	TTTAAATTTATTTTTTTTTTAAATTAATTAATTAATTAATTAATTAATTAATTAATTA	511

RESULT 10	
CNS021G2	
LOCUS	1146 bp DNA
DEFINITION	nigroviridis genome survey sequence T7 end of clone
	225004 of library G from Tetraodon nigroviridis, genomic survey
	sequence.
ACCESSION	AL176843
VERSION	AL176843.1 GI:7814900

KEYWORDS	GSS; genome survey sequence
SOURCE	Tetraodon nigroviridis.
ORGANISM	Tetraodon nigroviridis

REFERENCE
1 (pages 1 to 1146)
Roest-Crollius, H., Jallion, O., Dasilva, C., Bouneau, L., Fisher, C.,

TITLE	Human gene number estimate provided by genome wide analysis using Tetraodon nitoroviridis DNA sequence
1	1

REFERENCE	AUTHORS
2 (phases I to II46)	Roest-Crollius, H., Jallion, O., Dasilva, C., Filames, C., Fisher, C., Dallwitz, R., and Quilley, C.

TITLE Characterization and repeat analysis of the compact genome of the
Arabidopsis thaliana ecotype Col-0

REFERENCE 3 (bases 1 to 1146)

COMMENT
This sequence is a single read and was generated as part of a larger project.

<http://www.genoscope.cns.fr/Tetraodon>.
Location/DNA/144070

```
/organism="Tetraodon nigrroviridis"  
/db_xref="taxon:99883"
```

BASE COUNT	393 a	60 c	73 g	274 t	346 others
ORIGIN					

Query Match	5.8%;	Score 63.6;	DB 17;	Length 1146;
Best Local Similarity	31.8%;	Pred. No. 0.021;		
Matches 150;	Conservative 98;	Mismatches 222;	Indels 1;	Gaps

QY	534	ACTCTCGGAGTTCGCGAGCATTATCAAAAATTTATACAGAAAGGTTATTA	59
Db	591	ATTTAATTTRRRKTTAAKRRKKRAKTKRKTTHDMATTAARKKTTGRRKKRKT	65
QY	594	TTTAAAACTCAACACTGCATGAGATAAATTTCTTAAACACTTTTACGATGATTGTA	65
Db	651	KTTTAAGRRKKTKTTTANKTKAK-TAATAARKTTTARTTTTDTTTTAAKRRKKRKK	70
QY	654	CGATCTATCTAATGACTTTTTTTTCTTACACGCGTGATGAAAGTTATAGTACTATTAG	71
Db	710	AGKGRKRAKRRAKAKTKARRTTTTRAKATKARAARAAAAAAATTTATYKRRBKRRK	76
QY	714	CGAGACAAATGATTATAGATATATCCATTATCCATGATATTATGATATAAATAGCT	77
Db	770	KRRKKRAAARAAKKRRKTRGRKRRKGGARRKGGAAARGARRKKKGAAATTTTDD	82
QY	774	GTTAACTATTTTACATCGCAGCGTTCGCACATTTGTTTATTTATTAAGAGTTAT	83
Db	830	ATTATTTAATTTADATTTTADAKKKKKKKKKKTTTTTTTTTTTTTTTTTTTTTTTT	88
QY	834	AAATTTAAAGTATTTAAAGAGACATPACGAGCAACAAATGATTAAGMACGAGACAACAA	89
Db	890	TTTTTTTTTTTTTAAATTTTAAATTWAAAAATTTAAAAAAAATTTAAAAAAAATTTDAAKT	94
QY	894	AAAGCATCAAGCTCATGCGTTAGTTTAAAGCTTAATTAAGACATTTATTTAAATTTTAAG	95
Db	950	AAAATTTTTRKRDKRTTTTTTTTDRFTTWWTTTAAAAAAATTTWTATTTTATTTATTAAMA	100
QY	954	ACGATGATTAACATTAATTTTTCGACTCTTTTAAAAACCCCTCTACAA	1004

KEYWORDS GSS.
SOURCE Drosophila melanogaster.
ORGANISM Drosophila melanogaster
Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
Ephyridiidae; Drosophilidae; Drosophila.
REFERENCE 1 (bases 1 to 1101)
AUTHORS Genoscope.
TITLE Direct Submission
JOURNAL Submitted (02-JUN-1999) Genoscope - Centre National de Sequencage :
BP 191 91006 Evry cedex - FRANCE (E-mail : sequef@genoscope.cns.fr
- Web : www.genoscope.cns.fr)
COMMENT Determination of this BAC-end sequence was carried out as part of a
collaboration with the Berkeley Drosophila Genome Project (BDGP).
The BDGP is constructing a physical map of the Drosophila
melanogaster genome using these BACs. For further information
please see <http://www.fruitfly.org> The BDGP Drosophila
melanogaster BAC library was prepared by Kazutoyo Osoegawa and
Aaron Mammaster in Pieter de Jong's laboratory in the Department of
Cancer Genetics at the Roswell Park Cancer Institute in Buffalo,
NY. The library is named RPCI-98 and was constructed by partial
EcoRI digestion of Drosophila DNA provided by the BDGP from the
isogenic strain Y2; cn bw sp, the same strain used for the BDGP's
P1 and EST libraries. A more detailed description of the library
and how to order individual BAC clones, the entire library, or
filters for hybridization from the BACPAC Resource Center can be
found at http://bacpac.med.buffalo.edu/drosophila_bac.htm.
FEATURES
Source
1..1101
/organism="Drosophila melanogaster"
/db_xref="taxon:7227"
/clone="BACR24D09"
/clone_11b="RPCI-98"
/note="end : 77"
BASE COUNT 420 a 56 c 48 g 261 t 316 others
ORIGIN
Query Match 5.7%; Score 63.2; DB 17; Length 1101;
Best Local Similarity 41.1%; Pred. No. 0.025;
Matches 176; Conservative 51; Mismatches 199; Indels 2; Gaps 2
QY 558 TTATTCAGAAAATTTATGACGCAATGGGTTTATTAATTTAAAACTCAACCTTGATCAG 617
||| : : : : : ||| : : : : : ||| : : : : : ||| : : : : :
Db 1048 TTWTTATTTTATWATWATTAATATWATWATTTTAAAAAAATWAAATTTWTTTATTTAAAT 989
QY 618 ATAAATTTTCAAACTTTAGACGATGATTCGACGATCTATCTATGACTTTT 677
||| : : : : : ||| : : : : : ||| : : : : : ||| : : : : :
Db 988 TTAAATATTTTAAATTTTAAATTTTAAATTTTAAATTTTAAATTTTAAATTTTAAATTT 929
QY 678 TTTCACCAAGGTGATGAAAGTTTACTACTATTAGCAGACAGACAATTTGATTATAGATAT 737
||| : : : : : ||| : : : : : ||| : : : : : ||| : : : : :
Db 928 WATAAAAAAATATA-AAATATTTTAAATATTTWTTTAAATTAATTAATTTTATTTT 870
QY 738 ATCATTAATCCATGATATTTATGATTAATTAATAGCTGTTAAACATTTTCAGATCGCAGC 797
||| : : : : : ||| : : : : : ||| : : : : : ||| : : : : :
Db 869 WTTTATTTATTTTAAATTAATTTAAATWAAATTAATTTTATTTTATTTTATTTTATTTT 810
QY 798 TTTCGTGA-ACCTTTGTTTAAATTTTAAAGATTAAATTAATTAAGATTTAAAGGACA 856
||| : : : : : ||| : : : : : ||| : : : : : ||| : : : : :
Db 809 TTTTAAATTTTATTTTATTTTAAATTTTAAATTTTAAATTTTAAATTTTAAATTTTAAATTT 750
QY 857 TAACGAGCAACAAGTAATGAACACAGGAGAAACAAGCATGAAGCTATTTGTTAG 916
||| : : : : : ||| : : : : : ||| : : : : : ||| : : : : :
Db 749 AAAAAAAATWATAAAATAATTTTATTTTAAASVMTVMSSSCAATVTWAMA 690
QY 917 TTTAAGCTTAATAAGAGATTTTATTAATTTTATGACGATGATAACAATTAATTTTC 976
||| : : : : : ||| : : : : : ||| : : : : : ||| : : : : :
Db 689 AATTAATSTMAATSTVCASSSVWTTATATTTTAAATAAAMAKMAAATTTTSTGSTR 630
QY 977 TGACTTCT 984
||| : : : : : ||| : : : : : ||| : : : : : ||| : : : : :
Db 629 TTAATTTT 622

RESULT 13
CNS00587/c
LOCUS
DEFINITION
CNS00587 1009 bp DNA linear GSS 03-JUN-1999
Drosophila melanogaster genome survey sequence TET3 end of BAC #
BACR11N01 of RPCI-98 library from Drosophila melanogaster (fruit
fly), genomic survey sequence.
ACCESSION
AL057153
VERSION
AL057153.1
KEYWORDS
GI:4932541
SOURCE
GSS.
ORGANISM
Drosophila melanogaster.
Drosophila melanogaster.
Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
Ephydroidea; Drosophilidae; Drosophila.
REFERENCE
1 (bases 1 to 1009)
Genoscope.
AUTHORS
Direct Submission
TITLE
Submitted (02-JUN-1999) Genoscope - Centre National de Sequencage :
BP 191 91006 EVRY cedex - FRANCE (E-mail : seqref@genoscope.cns.fr
JOURNAL
- Web : www.genoscope.cns.fr)
COMMENT
Determination of this BAC-end sequence was carried out as part of a
collaboration with the Berkeley Drosophila Genome Project (BDGP).
The BDGP is constructing a physical map of the Drosophila
melanogaster genome using these BACs. For further information
please see <http://www.fruitfly.org> The BDGP Drosophila
melanogaster BAC library was prepared by Kazutoyo Osoegawa and
Aaron Mamoser in Pieter de Jong's laboratory in the Department of
Cancer Genetics at the Roswell Park Cancer Institute in Buffalo,
NY. The library is named RPCI-98 and was constructed by partial
EcoRI digestion of Drosophila DNA provided by the BDGP from the
isogenic strain y2; cn bw sp, the same strain used for the BDGP's
p1 and EST libraries. A more detailed description of the library
and how to order individual BAC clones, the entire library, or
filters for hybridization from the BACPAC Resource Center can be
found at http://bacpac.med.buffalo.edu/drosophila_bac.htm.
FEATURES
source
1..1009
/organism="Drosophila melanogaster"
/db_xref="taxon:7227"
/clone="BACR11N01"
/clone_1lb="RPCI-98"
/note="end : TET3"
BASE COUNT 263 a 116 c 96 g 477 t 57 others
ORIGIN
Query Match 5.7%; Score 62.8; DB 17; Length 1009;
Best Local Similarity 46.5%; Pred. No. 0.031;
Matches 148; Conservative 10; Mismatches 160; Indels 0; Gaps 0;
QY 1 AATCTCAAAATTTTAAATATTAGTCATATACAAATGCAAGAAATTCACAAAAATTT 60
DB 1006 AATATATAAANATWAAANATTAATTAATAAATTAATAAATTAATAAATTTTAAAT 947
QY 61 TGTAAACAGAACTCCAAATTTTTTTTTTTTGTGACAGAAATTAACAGATGAAGAAC 120
DB 946 NATAAAAAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 887
QY 121 TATTTGTTGTGAGTGAAGTAGTATATATACATTAAGCAAAATTTTAAAAATTAATAA 180
DB 886 AATTTAAATATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 827
QY 181 GCGTATACGCGCTCAAGATGTTATCTAGTAGGTGAATTAATTAATTAATTAATTAAT 240
DB 826 TAAAAAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 767
QY 241 CAGAAATTTGGACACAAATGAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 300
DB 766 AATTTATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 707
QY 301 TGAGTAAATGTGTTTCT 318
DB 706 TAAATTAATTAATTAATTTTWT 689

RESULT 14
CNS00591L
LOCUS
DEFINITION
CNS00591L 1101 bp DNA linear GSS 04-JUN-1999
Drosophila melanogaster genome survey sequence T7 end of BAC:
BACR29B23 of RPCI-98 library from Drosophila melanogaster (fruit
fly), genomic survey sequence.
ACCESSION
AL069706
VERSION
AL069706.1
KEYWORDS
GI:4949849
SOURCE
GSS.
ORGANISM
Drosophila melanogaster.
Drosophila melanogaster.
Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
Ephydroidea; Drosophilidae; Drosophila.
REFERENCE
1 (bases 1 to 1101)
Genoscope.
AUTHORS
Direct Submission
TITLE
Submitted (02-JUN-1999) Genoscope - Centre National de Sequencage :
BP 191 91006 EVRY cedex - FRANCE (E-mail : seqref@genoscope.cns.fr
JOURNAL
- Web : www.genoscope.cns.fr)
COMMENT
Determination of this BAC-end sequence was carried out as part of a
collaboration with the Berkeley Drosophila Genome Project (BDGP).
The BDGP is constructing a physical map of the Drosophila
melanogaster genome using these BACs. For further information
please see <http://www.fruitfly.org> The BDGP Drosophila
melanogaster BAC library was prepared by Kazutoyo Osoegawa and
Aaron Mamoser in Pieter de Jong's laboratory in the Department of
Cancer Genetics at the Roswell Park Cancer Institute in Buffalo,
NY. The library is named RPCI-98 and was constructed by partial
EcoRI digestion of Drosophila DNA provided by the BDGP from the
isogenic strain y2; cn bw sp, the same strain used for the BDGP's
p1 and EST libraries. A more detailed description of the library
and how to order individual BAC clones, the entire library, or
filters for hybridization from the BACPAC Resource Center can be
found at http://bacpac.med.buffalo.edu/drosophila_bac.htm.
FEATURES
source
1..1101
/organism="Drosophila melanogaster"
/db_xref="taxon:7227"
/clone="BACR29B23"
/clone_1lb="RPCI-98"
/note="end : T7"
BASE COUNT 419 a 91 c 60 g 299 t 232 others
ORIGIN
Query Match 5.7%; Score 62.6; DB 17; Length 1101;
Best Local Similarity 35.1%; Pred. No. 0.032;
Matches 140; Conservative 65; Mismatches 194; Indels 0; Gaps 0;
QY 8 AATTTATTAATTTTATGCTCAATACATGATGAAATTCACAAAAATTTTGTAAAC 67
DB 587 AATWTATTAATTAATTAATTAATTAATTAATTAATTTTAAATTAATTAATTAAT 646
QY 68 AGAACTTCCAAATTTTTTTTTTTTATGCAACAGAAATTAACAGATGAAGCAATTTTG 127
DB 647 TAAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 706
QY 128 TTGCGAATGGAAGTAGTATATACATTAAGCAAAATTTTAAAAATTAATAAACCCTATA 187
DB 707 ATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 766
QY 188 GCGCTCAAGTATGTTATCTAGTAGTGAATTAATTAATTAATTAATTAATTAATTAATTAAT 247
DB 767 WTTWTATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 826
QY 248 GGAACAAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 307
DB 827 WATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 886
QY 308 ATGTGTTTCTGCTATGAGGGCAAAAAAGCAATCCCAAAAGTCTACGGCTTGA 367

[illegible]

```
QY      246 TTGGGCAACATCGAAAGACGAATTAAAATTTAACCTTAAATAAATAAATTTGAGT 305
Db      711 ATATATAAAAAATAAAAAAAAAAATAAATAAATAAATAAATAAATAAATAAATTTAAAA
QY      306 AATATGTTCCTGACTATTCAGGGGCCAAAAAAGAACAATGCCAAA 353
Db      771 AAAAATAATTTTTTTTNAATTTTAATAAATAAATAAATAAATAAATAAATAA 818
```

Search completed: June 11, 2003, 06:58:35
Job time : 1673.1 secs

Search completed: June 11, 2003, 06:58:35
Job time : 1673.1 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: June 10, 2003, 22:00:04 ; Search time 953.95 Seconds
(Without alignments)

10921.753 Million cell updates/sec

Title: US-09-502-426b-1_COPY_6111_6468

Perfect score: 358

Sequence: 1 aaaaaaaaaagatgaagf.....tgcacaataccaacaacaag 358

Scoring table: IDENTITY NUC

Gapop 10.0 , Gapext 1.0

Searched: 2054640 seqs, 14551402878 residues

Total number of hits satisfying chosen parameters: 4109280

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : GenEmbl:*

1: gb_da:*

2: gb_htg:*

3: gb_in:*

4: gb_om:*

5: gb_ov:*

6: gb_pat:*

7: gb_ph:*

8: gb_pl:*

9: gb_pr:*

10: gb_ro:*

11: gb_sts:*

12: gb_sy:*

13: gb_un:*

14: gb_vi:*

15: em_ba:*

16: em_fun:*

17: em_hum:*

18: em_in:*

19: em_mu:*

20: em_om:*

21: em_or:*

22: em_ov:*

23: em_pat:*

24: em_ph:*

25: em_pl:*

26: em_ro:*

27: em_sts:*

28: em_un:*

29: em_vi:*

30: em_htg_hum:*

31: em_htg_inv:*

32: em_htg_other:*

33: em_htg_mus:*

34: em_htg_pln:*

35: em_htg_rtd:*

36: em_htg_mam:*

37: em_htg_vrt:*

38: em_sy:*

39: em_htgo_hum:*

40: em_htgo_mus:*

41: em_htgo_other:*

Pred. No. is the number of results predicted by chance to have a

score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	358	100.0	4818	8 AF044216	AF044216 Arabidops
2	356.4	99.6	84196	8 AIT345	AL132979 Arabidops
3	121	33.8	1691	8 AF412114	AF412114 Arabidops
4	58	16.2	5449	6 AX346543	AX346543 Sequence
5	58	16.2	7560	6 AX346125	AX346125 Sequence
6	57.2	16.0	173786	2 AC107582	AC107582 Rattus no
7	56.6	15.8	309233	2 AC098557	AC098557 Rattus no
8	56.4	15.8	5926	6 AX346554	AX346554 Sequence
9	55.6	15.5	12356	6 AX251264	AX251264 Sequence
10	55.6	15.5	169546	2 AC004157	AC004157 Plasmodu
11	55	15.4	7351	6 AX344930	AX344930 Sequence
12	55	15.4	20486	6 AX281500	AX281500 Sequence
13	54.8	15.3	93491	2 AC116967	AC116967 Dictyoste
14	54.8	15.3	349980	6 AX344570	AX344570 Sequence
15	54.6	15.3	123280	2 AC117076	AC117076 Dictyoste
16	54.6	15.3	268147	2 AC116966	AC116966 Dictyoste
17	54.2	15.1	2814	6 AX347029	AX347029 Sequence
18	54.2	15.1	6657	3 AC114263	AC114263 Dictyoste
19	54.2	15.1	67919	2 AC123610	AC123610 Mus muscu
20	54.2	15.1	204652	2 PFMA113P6	PFMA113P6
21	54	15.1	7138	6 AX323841	AX323841 Sequence
22	53.2	14.9	187013	2 AC116920	AC116920 Dictyoste
23	53.2	14.9	6047	2 AC115582	AC115582 Dictyoste
24	53.2	14.9	14147	6 AX251501	AX251501 Sequence
25	53.2	14.9	14147	6 AX347392	AX347392 Sequence
26	53.2	14.9	14147	6 AX349113	AX349113 Sequence
27	53	14.8	7261	6 AX251424	AX251424 Sequence
28	53	14.8	133501	2 AC116956	AC116956 Dictyoste
29	53	14.8	349980	6 AX346040	AX346040 Sequence
30	52.8	14.7	5678	6 AX316979	AX316979 Dictyoste
31	52.8	14.7	107739	2 AC005139	AC005139 Plasmodu
32	52.8	14.7	256172	2 AC005139	AC005139 Plasmodu
33	52.8	14.7	310779	2 AC005140	AC005140 Plasmodu
34	52.6	14.7	6657	3 AC114263	AC114263 Dictyoste
35	52.6	14.7	24091	2 AC014610	AC014610 Drosophi
36	52.6	14.7	106993	3 AE002751	AE002751 Drosophi
37	52.6	14.7	349980	6 AX344564	AX344564 Sequence
38	52.4	14.6	150903	2 AC113218	AC113218 Rattus no
39	52.2	14.6	6418	6 AX251768	AX251768 Sequence
40	52.2	14.6	6418	6 AX345224	AX345224 Sequence
41	52.2	14.6	16750	6 AX251068	AX251068 Sequence
42	52.2	14.6	16750	6 AX345423	AX345423 Sequence
43	52.2	14.6	66441	3 PFMA11P4	PFMA11P4
44	52.2	14.6	234112	3 PFMA14P2	PFMA14P2
45	51.8	14.5	627	11 CNS06EWH	AL395655 T7 end of

ALIGNMENTS

RESULT 1

AF044216 4818 bp DNA linear PLN 25-JUN-2001

LOCUS Arabidopsis thaliana sterold 22-alpha-hydroxylase (DMF4) gene,

DEFINITION complete cds.

ACCESSION AF044216

VERSION AF044216.1 GI:2935341

KEYWORDS

SOURCE Arabidopsis thaliana

ORGANISM Arabidopsis thaliana

REFERENCE 1 (bases 1 to 4818)

Choe,S., Dilkes,B.P., Fujioaka,S., Takatsuto,S., Sakurai,A. and

TITLE Feldmann, K. A.
 The DWF4 gene of Arabidopsis encodes a cytochrome P450 that
 mediates multiple 22alpha-hydroxylation steps in brassinosteroid
 biosynthesis
 JOURNAL Plant Cell 10 (2), 231-243 (1998)
 MEDLINE 98158690
 PUBMED 9490746
 REFERENCE 2 (bases 1 to 4818)
 AUTHORS Choe, S., Dilkes, B. P., Azpilroz, R. and Feldmann, K. A.
 JOURNAL Direct Submission
 TITLE Submitted (22-JAN-1998) Plant Sciences, University of Arizona,
 Tucson, AZ 85721, USA
 FEATURES Location/Qualifiers
 source 1. .4818
 /organism="Arabidopsis thaliana"
 /strain="Ws-2"
 /db_xref="taxon:3702"
 /chromosome="III"
 /map="8 cM from nga6"
 990. .4398
 /gene="DWF4"
 /note="DWF4"
 /allele="dwa4-1; dwa4-2; dwa4-3; dwa4-4"
 990. .1055
 /gene="DWF4"
 /join("1133. .1353,1434. .1758,1844. .1996,2095. .2409,
 2562. .2654,2746. .2824,2931. .3040,3795. .4398)
 /product="steroid 22-alpha-hydroxylase"
 /gene="DWF4"
 /join("1133. .1353,1434. .1758,1844. .1996,2095. .2409,
 2562. .2654,2746. .2824,2931. .3040,3795. .4040)
 /note="DWF4"
 /function="brassinosteroid biosynthetic enzyme"
 /note="member of the cytochrome P450 superfamily; CYP90B1"
 /codon_start=1
 /product="steroid 22-alpha-hydroxylase"
 /protein_id="AAC05093.1"
 /db_xref="GI:2935342"
 /translation="MEETEHTLLPLLLPSLLSLFLILKRRNRKTRNDLPKGS
 GMPLEETIGYLRPYATLGDPMQOHVSXGKTYSNLGEPTIVSADGLRFTIO
 NEGRLEFCSYPSIGGILGKMSMLVGVDMHDMRMSISNLFSLHARLRTLLDVERH
 TLEVLDSMOONSIFSAODEAKKFTFIRMAKHMNDGSEETOLKKEYTEFKGVYS
 APLNPGTAYHKLQSRATLTKFERKMERKDIKEDEDEEVEVTEDEAKSKSH
 VKKQRTDDDLGLVHLKHSNSTQIIDLILSLPAGETSSVALALIFFLQACPAV
 EELREHLEIARAKKELGSELNMDYKKMDFQCVINETRLGNVVRFLRRALDY
 RYKGYDIPSGMKVLPVISAHLDNSRYDQNLFPNPMQONNAGSSGSGSFTGMN
 NYMPFGGPRLCAGSELAKLEMAVFIHHLVILKFWELAEDDQFAPFVDFPGLPIR
 VSRIL"
 BASE COUNT 1566 a 727 c 888 g 1637 t
 ORIGIN
 Query Match 100.0%; Score 358; DB 8; Length 4818;
 Best Local Similarity 100.0%; Pred. No. 5e-51;
 Matches 358; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 1 AAAAAAAAAAGATGAAATTTTATTTCTCTCTTTTGTGATAATTTAAATCA 60
 DB 4041 AAAAAAAAAAGATGAAATTTTATTTCTCTCTTTTGTGATAATTTAAATCA 4100
 QY TTTTGTGGCCATGATATATAAATTGATAATATATATATGATATTCGTTT 120
 DB 4101 TTTTGTGGCCATGATATATAAATTGATAATATATATATGATATTCGTTT 4160
 QY TTAGTTCGGGTTTGAAGAAAGGTTTCGACTTCGAAAGTGAGATGATATAGATTGG 180
 DB 4161 TTAGTTCGGGTTTGAAGAAAGGTTTCGACTTCGAAAGTGAGATGATATAGATTGG 4220
 QY TTAGTTCGGGTTTGAAGAAAGGTTTCGACTTCGAAAGTGAGATGATATAGATTGG 180
 DB 181 GAGCTGAGTTGAGCTTTTGACATTTTGATGATGATGTTGATATATAGTGCACACT 240
 DB 4221 GAGCTGAGTTGAGCTTTTGACATTTTGATGATGATGTTGATATATAGTGCACACT 4280
 QY ATTAACCTTAATGGCTTCTATTAAGGCCCAATATATATAGATTATACAAATGAC 300
 DB 4281 ATTAACCTTAATGGCTTCTATTAAGGCCCAATATATATAGATTATACAAATGAC 4340
 QY 301 AACTTTACTTCGTTTGTGATCCGACCAATTAACAAATGTGCAATACCAACACAG 358
 DB 4341 AACTTTACTTCGTTTGTGATCCGACCAATTAACAAATGTGCAATACCAACACAG 4398
 RESULT 2
 ATT3A5
 LOCUS
 DEFINITION Arabidopsis thaliana DNA chromosome 3, BAC clone T3A5.
 ACCESSION AL132979
 VERSION AL132979.2 GI:6782244
 KEYWORDS
 SOURCE
 ORGANISM Arabidopsis thaliana.
 Arabidopsis thaliana.
 Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
 Rosidae; eustosids II; Brassicales; Brassicaceae; Arabidopsis.
 1 (bases 1 to 84196)
 Bloembergen, H., Mewes, H.W., Lemcke, K., Mayer, K.F.X., Quetier, F. and
 Salanoubat, M.
 JOURNAL Unpublished
 REFERENCE 2 (bases 1 to 84196)
 AUTHORS EU Arabidopsis sequencing project.
 JOURNAL Direct Submission
 TITLE Submitted (25-JAN-2000) MIPS, at the Max-Planck-Institut fuer
 Biochemie, Am Klopferspitz 18a, D-82152 Martinsried, FRG. E-mail:
 bloem@mips.biochem.mpg.de, mayer@mips.biochem.mpg.de, Project
 Coordinator: Marcel Salanoubat and Francis Quetier, Groupement
 d'Interet Public, Centre National de Sequencage - GENOSCOPE, 2 rue
 Gaston Cremieux, BP191, 91006 Evry Cedex, France;
 http://www.genoscope.cns.fr
 On Jan 27, 2000 this sequence version replaced gi:6434247.
 Information on performance of analysis and a more detailed
 annotation of this entry and other sequences of chromosomes 3, 4
 and 5 can be viewed at: http://www.mips.biochem.mpg.de/proj/thal/
 FEATURES Location/Qualifiers
 source 1. .84196
 /organism="Arabidopsis thaliana"
 /variety="Columbia"
 /db_xref="taxon:3702"
 /chromosome="3"
 2963. .3338
 /note="355 bp LTR"
 3339. .3564
 /note="LTR-transposon"
 7565. .7922
 /note="367bp LTR"
 12339. .13197
 /gene="T3A5.10"
 /join("12339. .12665,12742. .12803,12888. .13055,13125. .13197)
 /gene="T3A5.10"
 /note="similarity to cyclin-dependent kinase inhibitor
 (ICK1), Arabidopsis thaliana, EMBL:AT094772.1"
 /codon_start=1
 /product="putative protein"
 /protein_id="CAB62432.1"
 /db_xref="GI:6561966"
 /translation="MAVRRRRRDYVENCVTITTVKRRKMEERVDLVESRIITSPCV
 QATNRGIVARNAGASEISVYVRRDSPVEVQCIIEEDSSVSCSSEKSKRR
 IERVLDENNDDETFETSNITVDLNSSESMNDDSSVAVEDVESRRRLKSLHETV.
 KEALEDPEFQVAKDRNLKLLKCSMKNFDEKDEPLGGGRYEMVKNLP"
 12339. .12665
 /gene="T3A5.10"
 /number=1
 12666. .12741
 /gene="T3A5.10"
 /number=1
 12742. .12803
 /gene="T3A5.10"
 /number=2
 12804. .12887
 Intron
 exon
 Intron

exon /gene="T3A5.10"
/number=2
12888. .13055
/gene="T3A5.10"
/number=3
13056. .13124
/gene="T3A5.10"
/number=3
13125. .13197
/gene="T3A5.10"
/number=4
13588. .16470
/gene="T3A5.20"
complement(join(15858. .15901,16014. .16470))
/note="similarity to predicted proteins, Arabidopsis
thaliana"
/codon_start=1
/product="putative protein"
/protein_id="CAB62433.1"
/db_xref="GI:6561967"
/translation="MARLEMNNGAARISFSENEFEIRSEKSNKSNINSRSPS
MSPAFSAVDYDSMIPDADEFIKGKILPEKETSHVHRTTGELLTTEEGSDMGNF
SLRPLTSSSSSTKGTRELLGKRTHVRSKTKDKVNEEVLSDHKITISGVATRECC
QVADLR"
exon complement(15858. .15901)
/number=1
complement(15902. .16013)
/number=1
complement(16014. .16470)
/number=2
18303. .19931
/gene="T3A5.30"
complement(18303. .19931)
/codon_start=1
/product="scartecrow-like 7 (SCL7)"
/protein_id="CAB62434.1"
/db_xref="GI:6561968"
/translation="MAVYCTDNGNLMATAOQLIKOKOOQOOSHOQOEEBOEPPMPN
PSFGTLLSGSGDFEVOYNDPGFPHLEHNOAAVASEFSDDEMESIINGDAS
GNPEPPEYDGFVSPSPRSAPSYLNRYKVDASOQLPPASPAIASSPSPSQ
HPPPEPPDDINDQPIKAIHDYARREPTDILRIKESVSSGDTIOVGTFAE
ALSKEETSPSSSSSSSLEDFTLSTKTLNDACPSKFAHLTANALILEANOSNNIHI
VDFGFQGIOWSALLQALATNGSSGKPRIRISGIPASISGSPSILANGRIJRDPA
ALINDNEFEYPLTPIQLNGSSGKPRIRISGIPASISGSPSILANGRIJRDPA
LSLRINRIYLTGSEYNSLRVFEKARKKSLRYSAAVESLEPNLDRSKRIAREVLT
RGRITMDLVISDQNNKTRFGRLMEKEKQMRVMEKAGPEPKPSNAYSAQKLIM
NNVNTSLVSESGEFTSLANNVPLLTIVSMR"
exon complement(18303. .19931)
/number=1
26093. .28999
/gene="T3A5.40"
complement(join(26093. .26338,27091. .27200,27308. .27386,
27478. .27570,27723. .28037,28136. .28288,28374. .28698,
28779. .28999))
/note="Contains cytochrome P450 cysteine heme-iron ligand
signature AA455-464:Prokaryotic membrane lipoprotein lipid
attachment site AA320-330"
/codon_start=1
/product="steroid 22-alpha-hydroxylase (DWF4)"
/protein_id="CAB62435.1"
/db_xref="GI:6561969"
/translation="MFEETHTTLPRLPRLSLTLLFTLLKRRNRKTRPNLPBGKS
GMPLFSLTIGYIKPRTATTDGDMQOHVSKYGRISNLNGEPTIYADAGNRPITQ
NEGRIFECYSRSTIGILGKSMYLIVGDHNRDRSISLNFLSHARITLLDOVERH
TLPLVLDNQOQSIFSAODEAKKFTFNMAKIMSDGEDEETDQKREYTFPKGSVS
ALPNLPGTAHKAALQSRATTLIKFTIERKMEERKLDIKEDQDEEYKTEDAEKSSSH
VRKORTDDLLGVWLKSHNSTEODILDLISILFAGHETSSVAIALAIPLOACPAY
EALREHREHILARRAKKEGESELDNMDYQKMDPFCVNETLIRGVNREPLHRAALDY
RYKGDIDSGKAVLPVISAHLNDRDQPLTEPMDMOQOONNASSSGSGSSTGN
NMPMGCGPRCCASSELAKLEMAVFIHLHLVLFKNWELAEEDKPPAFPFVDFPGLPLR
VSRIL"
exon complement(26093. .26338)
/number=1

Query Match	Best Local Similarity	Score	Length
DB 26092	99.7%	356.4	84196
Matches 357;	Conservative 0;	Mismatches 1;	Indels 0; Gaps 0;

Query Match	Best Local Similarity	Score	Length
DB 26092	99.7%	356.4	84196
Matches 357;	Conservative 0;	Mismatches 1;	Indels 0; Gaps 0;

```

|||||
Db 26032 TTTTTCCTGGCCATGATATATAAAATTGGATTAATATATATGATATTCGTTT 25973
|||
Qy 121 TTAGTCGGGTTGAGAAAAGGTTTCGACCTTCGAAAGTGCAGATGATATATGATTGG 180
|||
Db 25972 TTAGTCGGGTTGAGAAAAGGTTTCGACCTTCGAAAGTGCAGATGATATATGATTGG 25913
|||
Qy 181 GACCTAGGTTGAGTCTTTGACATTTGATTTGATGATGTTGATTTAGTGTGACACT 240
|||
Db 25912 GACCTAGGTTGAGTCTTTGACATTTGATTTGATGATGTTGATTTAGTGTGACACT 25853
|||
Qy 241 ATTAACCTTAATGCGCTTTCTATAGGCCCAATTTATAGCATTTATACAAAGTGAC 300
|||
Db 25852 ATTAACCTTAATGCGCTTTCTATAGGCCCAATTTATAGCATTTATACAAAGTGAC 25793
|||
Qy 301 AACTTTACTCTGTTTTCGATCCGAAGCATTAACAATTCGCAATACCAACGACAG 358
|||
Db 25792 AACTTTACTCTGTTTTCGATCCGAAGCATTAACAATTCGCAATACCAACGACAG 25735
|||

RESULT 3
AF412114 1691 bp mRNA linear PLN 24-SEP-2001
LOCUS Arabidopsis thaliana AT3g50660/T3A5_40 mRNA, complete cds.
DEFINITION AF412114
ACCESSION AF412114.1 GI:15724347
VERSION FLI.CDNA.
KEYWORDS Arabidopsis thaliana
SOURCE Arabidopsis thaliana
ORGANISM Arabidopsis thaliana
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
Rosidae; eurosids II; Brassicales; Brassicaceae; Arabidopsis;
1 (bases 1 to 1691)
REFERENCE
AUTHORS Cheuk,R., Chen,H., Kim,C.-J., Koesema,E., Meyers,M.C., Banh,J.,
Bowler,L., Carlincl,P., Dale,J.M., Goldsmith,A.D., Hayashizaki,Y.,
Ishida,J., Jiang,P.X., Jones,T., Kamiya,A., Karlin-Neumann,G.,
Kawai,J., Lam,B., Lee,J.M., Lin,J., Liu,S.X., Miranda,M.,
Narusaka,M., Nguyen,M., Onodera,C.S., Palm,C.J., Pham,P.K.,
Quach,H.L., Sakurai,T., Satou,M., Seki,M., Southwick,A., Tang,C.C.,
Toriumi,M., Yamada,K., Yamamura,Y., Yu,G., Yu,S., Shinozaki,K.,
Davis,R.W., Theologis,A. and Ecker,J.R.
Arabidopsis cDNA clones
Unpublished
2 (bases 1 to 1691)
Cheuk,R., Chen,H., Kim,C.-J., Koesema,E., Meyers,M.C., Banh,J.,
Bowler,L., Carlincl,P., Dale,J.M., Goldsmith,A.D., Hayashizaki,Y.,
Ishida,J., Jiang,P.X., Jones,T., Kamiya,A., Karlin-Neumann,G.,
Kawai,J., Lam,B., Lee,J.M., Lin,J., Liu,S.X., Miranda,M.,
Narusaka,M., Nguyen,M., Onodera,C.S., Palm,C.J., Pham,P.K.,
Quach,H.L., Sakurai,T., Satou,M., Seki,M., Southwick,A., Tang,C.C.,
Toriumi,M., Yamada,K., Yamamura,Y., Yu,G., Yu,S., Shinozaki,K.,
Davis,R.W., Theologis,A. and Ecker,J.R.
Direct Submission
Submitted (21-AUG-2001) Salk Institute Genomic Analysis Laboratory
(SIGNAL), Plant Biology Laboratory, The Salk Institute for
Biological Studies, 10010 N. Torrey Pines Road, La Jolla, CA 92037,
USA
RIKEN Genomic Sciences Center (GSC) members carried out the
sequencing and annotation of the RAFL cDNAs: Cheuk,R., Chen,H.,
Kim,C.J., Koesema,E., Meyers,M.C., Shinn,P., Banh,J., Bowser,L.,
Dale,J.M., Goldsmith,A.D., Jiang,P.X., Jones,T., Kamiya,A., Karlin-Neumann,G.,
Lam,B., Lee,J.M., Lin,J., Liu,S.X., Miranda,M., Nguyen,M.,
Onodera,C.S., Palm,C.J., Pham,P.K., Quach,H.L., Southwick,A.,
Tang,C.C., Toriumi,M., Yamada,K., Yamamura,Y., Yu,G., Yu,S.,
Davis,R.W., Theologis,A., and Ecker,J.R.
Hayashizaki,Y. and Shinozaki,K.

COMMENT
The Salk, Stanford, PECC (SSP) Consortium members carried out the
sequencing and annotation of the RAFL cDNAs: Cheuk,R., Chen,H.,
Kim,C.J., Koesema,E., Meyers,M.C., Shinn,P., Banh,J., Bowser,L.,
Dale,J.M., Goldsmith,A.D., Jiang,P.X., Jones,T., Kamiya,A., Karlin-Neumann,G.,
Lam,B., Lee,J.M., Lin,J., Liu,S.X., Miranda,M., Nguyen,M.,
Onodera,C.S., Palm,C.J., Pham,P.K., Quach,H.L., Southwick,A.,
Tang,C.C., Toriumi,M., Yamada,K., Yamamura,Y., Yu,G., Yu,S.,
Davis,R.W., Theologis,A., and Ecker,J.R.
Cheuk,R. (SSP/Salk) and Seki,M. (RIKEN GSC) contributed equally to

```

```

this work. Shinozaki,K. (RIKEN GSC) and Ecker,J.R. (SSP/Salk)
contributed equally to this work as PIs.
FEATURES
source
location/Qualifiers
1. 1691
/organism="Arabidopsis thaliana"
/db_xref="taxon:3702"
/chromosome="3"
/clone="RAFL06-76-F19(R13551)"
/note="ecotype: Columbia"
1..28
5'/UTR
CDS
29..1570
/note="steroid 22-alpha-hydroxylase (DWF4)"
/codon_start=1
/product="AT3g50660/T3A5_40"
/protein_id="AAL06567.1"
/db_xref="GI:15724348"
/translation="MFEFHHLLPLLLPSLISLFLILKRRNRKRPNPGRKS
GMPPLGETITGLKPYRTTITGDFRQGVSKRTKRYRNLREPTIVADAGLNFTILQ
NEGRLEFCSYPRISIGILGRWSMLVLGDHMRDRSISLNLSHARKLTILKQVNH
TLRLVDSMOONSIFSAQDEAKKFTFNLMARHMSMDGEETDLKREYVFMGVS
APLNPCTAYHKLQSRATILKFERMEERKIDKEDEEVEKTEDEAKMSKH
VRKORTDDDLGWYLKHSNSTEOILDLILSLPAGHETSVAALAIFFLOACPKAY
ELRBEHLFLARAKKEIGESLEMDVKKKMDPQCIVNLTIRLGNVRLPHRKALKYV
RYKGYDLPSSMKVLPIYSAYHLDNSRDQENLRPRMWOONNGASSGSGSESTWKN
NMFEGGPRLCAGSELAKLEMAVFHHLVLRNWELEDDKPPAFPFVDFPGLPLR
VSRIL"
1571..1691
3'UTR
BASE COUNT 524 a 286 c 363 g 518 t
ORIGIN
Query Match 33.8% Score 121; DB 8; Length 1691;
Best Local Similarity 100.0%; Pred. No. 2,7e-11;
Matches 121; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 AAAAAAAAAAAGATGATATATAAAATTGGATTAATATATGATATTCGTTT 60
Db 1571 AAAAAAAAAAAGATGATATATAAAATTGGATTAATATATGATATTCGTTT 1630
|||
Qy 61 TTTTTCCTGGCCATGATATATAAAATTGGATTAATATATGATATTCGTTT 120
|||
Db 1631 TTTTTCCTGGCCATGATATATAAAATTGGATTAATATATGATATTCGTTT 1690
|||
Qy 121 T 121
Db 1691 T 1691

RESULT 4
AX346543 5449 bp DNA linear PAT 01-FEB-2002
LOCUS AX346543
DEFINITION Sequence 1614 from Patent WO0200928.
ACCESSION AX346543
VERSION AX346543.1 GI:18494429
KEYWORDS
SOURCE synthetic construct.
ORGANISM synthetic construct
artificial sequences.
1
REFERENCE
AUTHORS Olek,A., Piepenbrock,C. and Beiln,K.
TITLE Diagnosis of diseases associated with the immune system
JOURNAL Patent: WO 0200928-A 1614 03-JAN-2002;
Epidemiol Infect 124: 1-11
FEATURES
source
location/Qualifiers
1. 5449
/organism="synthetic construct"
/db_xref="taxon:32630"
/note="chemically treated genomic DNA (Homo sapiens)"
BASE COUNT 1146 a 90 c 1299 g 2914 t
ORIGIN
Query Match 16.2% Score 58; DB 6; Length 5449;
Best Local Similarity 65.4%; Pred. No. 0.68;
Matches 85; Conservative 0; Mismatches 45; Indels 0; Gaps 0;

```

QY 4 AAAAAAAAAAGATGAACTATTTTATCTCTCTTTTGTGATATTTAAATCATTT 63
 DB 5256 AATTAATAAATTAAGATATTTTGTGATATTTTGTGATATTTTAAATGATT 5315
 QY 64 TTTTCCCAATGATATTAATAAATTTGATTAATATATGATATTCGTTTGA 123
 DB 5316 TGTATTAAGTTTATTAATATATTTTATATATTTTATATTTGATATTTTATTT 5375
 QY 124 GTTCGGGTTT 133
 DB 5376 TTTTGGGTTT 5385

RESULT 5
AX346125

LOCUS AX346125 7560 bp DNA linear PAT 01-FEB-2002
 DEFINITION Sequence 1196 from Patent WO0200928.
 ACCESSION AX346125
 VERSION AX346125.1 GI:18494011
 KEYWORDS

SOURCE
 ORGANISM
 REFERENCE
 1
 AUTHORS
 TITLE
 JOURNAL
 Epigenomics AG (DE)
 Location/Qualifiers
 1..7560
 /organism="synthetic construct"
 /db_xref="taxon:32630"
 /note="chemically treated genomic DNA (Homo sapiens)"

BASE COUNT 2154 a 47 c 1534 g 3805 t
 ORIGIN

Query Match 16.2%; Score 58; DB 6; Length 7560;
 Best Local Similarity 51.6%; Pred. No. 0.63;
 Matches 159; Conservative 0; Mismatches 145; Indels 4; Gaps 1;

QY 2 AAAAAAAAAAGATGAAAGTATTTTATCTCTCTTTTGTGATATTTAAATCAT 61
 DB 2728 AATAAATTAAGATATTTTATTTTGTGATATTTTAAATGATT 2787
 QY 62 TTTTCCCAATGATATTAATAAATTTGATTAATATATGATATTCGTTTGA 121
 DB 2788 TTTTATTCGGAATTTTAAATTTTAAAGATATTTTATTTATGTCGTTATT 2847
 QY 122 TACTTGGGTTTGAGAAAGGTTTCGACTTCGAAAGTGACGATATAGATTGG 181
 DB 2848 TGTATTAAGTTTAAATATTTTATTTTATTTTATTTTAAATAAATTAAT 2907
 QY 182 AGCTAGCTGAGTCTTGACATTTTGTATGATGT---TGTGATTAATAGTGCAC 237
 DB 2908 TATTAATAAATTAATTTGATTAATATGATTAATAATGATGATGATTTTAAATTT 2967
 QY 238 ACTATTAACCTTAATGGCTTCTATAGCCCAATTAATTCGATTAACAAAGT 297
 DB 2968 AATTTTGAATTAATAATTTTATTTTAAAGAAATTAATTAATTAATAAGA 3027
 QY 298 GACACTT 305
 DB 3028 GATTAATT 3035

KEYWORDS
SOURCE
ORGANISMREFERENCE
AUTHORS

HTG: HTGS, PHASE1.
 Norway rat.
 Rattus norvegicus
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;
 Rattus.
 1 (bases 1 to 173786)

Muzny,D.M., Adams,C., Adio-Oduola,B., Ali-osman,F.R., Allen,C.,
 Alshrooks,S.L., Amaratunga,H.C., Aye,J.R., Ayele,M., Banks,T.,
 Barbara,J., Benton,J., Bimge,K., Blankenburg,K., Bonin,D.,
 Bouck,J., Bowe,S., Brieva,M., Brown,E., Brown,M., Bryant,N.P.,
 Buhay,C., Burch,P., Burkett,C., Burrell,K.L., Byrd,N.C.,
 Caron,T.F., Carter,M., Cavazos,S.R., Checko,J., Chavez,D.,
 Chen,G., Chen,R., Chen,Z., Chowdhry,I., Christopoulos,C.,
 Cleveland,C.D., Cox,C., Coyle,M.D., Dathorne,S.R., David,R.,
 Davila,M.L., Davis,C., Davy-Carroll,L., Dederich,D.A.,
 Delaney,K.R., Delgado,O., Denn,A.L., Ding,Y., Dinh,H.H.,
 Douthwaite,K.J., Draper,H., Dugan-Rocha,S., Durbin,K.J.,
 Earnhart,C., Edgar,D., Edwards,C.C., Elhaj,C., Escotto,M.,
 Falls,T., Ferraguto,D., Flagg,N., Ford,J., Foster,P., Frantz,P.,
 Gabisi,A., Gao,J., Garcia,A., Garner,T., Gatz,N., Gill,R.,
 Gorell,J.H., Guevara,M., Gunatane,P., Hale,S., Hamilton,K.,
 Harris,C., Harris,K., Hart,M., Haylak,P., Hawes,A., Hernandez,J.,
 Hernandez,O., Hodgson,A., Hognes,M., Holloway,C., Hollins,B.,
 Homsi,F., Howard,S., Huber,J., Huijck,S., Hume,J., Jackson,L.E.,
 Jacobson,B., Jia,Y., Johnson,R., Jolivet,S., Joudah,S.,
 Karlsson,E., Kelly,S., Khan,U., King,L., Kovach,J., Kovar,C.,
 Karlovic,J., Kureshi,A., Landry,N., Leal,B., Lewis,L.C., Lewis,L.,
 Li,J., Li,Z., Lichtarge,O., Lieu,C., Liu,J., Liu,Y., Louised,H.,
 Lozano,R.J., Lu,X., Lucier,A., Lucier,R., Luna,R., Ma,J.,
 Maheshwari,M., Mapua,P., Martin,R., Martindale,A., Martinez,E.,
 Massey,E., MaWhiney,E., McLeod,M.P., Meador,M., Mei,G., Metzger,M.,
 Miner,G., Miner,Z., Mitchell,T., Mohabbat,K., Morgan,M., Morris,S.,
 Moser,M., Neal,D., Newton,J., Newton,J., Nguyen,A., Nguyen,N.,
 Nguyen,N., Nickerson,E., Nwokkenko,S., Ogun,N., Okunodu,G.,
 Oragunye,N., Oviado,R., Pace,A., Payton,B., Peery,J., Perez,L.,
 Peters,L., Pickens,R., Primus,E., Pu,L.L., Quiles,M., Ren,Y.,
 Rivers,M., Rojas,A., Rojubokeyan,I., Rolfe,M., Ruiz,S., Savary,G.,
 Scherren,S., Scott,G., Shen,H., Shoochthai,N., Sisson,I.,
 Sodergren,E., Sonalke,T., Sparks,A., Stanley,H., Stone,H.,
 Sutton,A., Syatel,C., Taylor,T., Telford,B., Thomas,N., Thomas,S.,
 Tansey,J., Taylor,C., Taylor,T., Telford,B., Thomas,N., Thomas,S.,
 Wang,S., Vasquez,L., Vera,Y., Villalobos,D., Vinson,R., Wang,Q.,
 Williams,G., Williamson,A., Wleczek,R., Wooden,S., Worley,K.,
 Wu,C., Wu,Y., Wu,Y.F., Zhou,J., Zorrilla,S., Nelson,D.,
 Weinstein,G. and Gibbs,R.
 Direct Submission
 Unpublished
 2 (bases 1 to 173786)
 Worley,K.C.
 Direct Submission
 Submitted (23-JAN-2002) Human Genome Sequencing Center, Department
 of Molecular and Human Genetics, Baylor College of Medicine, One
 Baylor Plaza, Houston, TX 77030, USA
 3 (bases 1 to 173786)
 Worley,K.C.
 Direct Submission
 Submitted (13-JUL-2002) Human Genome Sequencing Center, Department
 of Molecular and Human Genetics, Baylor College of Medicine, One
 Baylor Plaza, Houston, TX 77030, USA
 On Jul 12, 2002 this sequence version replaced gi:18846355.

TITLE
JOURNAL
REFERENCE
AUTHORS
TITLE
JOURNALREFERENCE
AUTHORS
TITLE
JOURNAL

COMMENT

Center: Baylor College of Medicine
 Center code: BCM
 Web site: http://www.hgsc.bcm.tmc.edu/
 Contact: hgsc-help@bcm.tmc.edu

Project Information
 Center project name: GNMJF

Sequencing statistics
 Summary Statistics
 Sequencing vector: Plasmid

Chemistry: Dye-terminator Big Dye: 100% of reads

RESULT 6
 AC107582 173786 bp DNA linear HTG 13-JUL-2002
 LOCUS Rattus norvegicus clone CH230-29B17, *** SEQUENCING IN PROGRESS
 DEFINITION *** 97 unordered pieces.
 ACCESSION AC107582
 VERSION AC107582.3 GI:21736513

Assembly program: Phrap; version 0.990329
 Consensus quality: 85701 bases at least Q40
 Consensus quality: 94901 bases at least Q30
 Consensus quality: 103067 bases at least Q20

NOTE: Estimated insert size may differ from sequence length
 (see http://www.hgsc.bcm.tmc.edu/docs/genbank_draft_data.html).
 NOTE: This is a 'working draft' sequence. It currently
 consists of 92 contigs. The true order of the pieces
 is not known and their order in this sequence record is
 arbitrary. Gaps between the contigs are represented as
 runs of N, but the exact sizes of the gaps are unknown.
 This record will be updated with the finished sequence
 as soon as it is available and the accession number will
 be preserved.

1 1044: contig of 1044 bp in length
 1144: gap of unknown length
 1145: contig of 1120 bp in length
 2265: gap of unknown length
 2265: contig of 1286 bp in length
 3651: gap of unknown length
 3751: contig of 1177 bp in length
 4928: gap of unknown length
 5027: contig of 1163 bp in length
 6191: gap of unknown length
 6290: gap of unknown length
 6291: contig of 1016 bp in length
 7307: gap of unknown length
 7407: contig of 1083 bp in length
 8490: gap of unknown length
 8589: contig of 1189 bp in length
 8590: gap of unknown length
 9779: contig of 1305 bp in length
 9879: gap of unknown length
 11183: contig of 1105 bp in length
 11283: gap of unknown length
 11284: contig of 1105 bp in length
 12389: gap of unknown length
 12488: contig of 1055 bp in length
 13543: gap of unknown length
 13643: contig of 1160 bp in length
 13644: gap of unknown length
 14803: contig of 1106 bp in length
 14804: gap of unknown length
 14904: contig of 1106 bp in length
 16010: gap of unknown length
 16109: contig of 1201 bp in length
 17310: gap of unknown length
 17311: contig of 1006 bp in length
 17411: gap of unknown length
 18416: contig of 1229 bp in length
 18417: gap of unknown length
 18517: contig of 1229 bp in length
 19746: gap of unknown length
 19845: contig of 1100 bp in length
 20946: gap of unknown length
 21046: contig of 1087 bp in length
 22132: gap of unknown length
 22232: contig of 1264 bp in length
 22233: gap of unknown length
 23497: contig of 1132 bp in length
 23597: gap of unknown length
 24728: contig of 1154 bp in length
 24828: gap of unknown length
 25982: contig of 1130 bp in length
 25983: gap of unknown length
 26082: contig of 1130 bp in length
 27212: gap of unknown length
 27213: contig of 1224 bp in length
 28536: gap of unknown length
 28636: contig of 1229 bp in length
 28637: gap of unknown length
 29865: contig of 1150 bp in length
 29866: gap of unknown length
 31115: contig of 1096 bp in length
 31116: gap of unknown length
 32312: contig of 1049 bp in length
 32313: gap of unknown length
 33460: contig of 1199 bp in length
 33461: gap of unknown length
 34760: contig of 1078 bp in length
 34860: gap of unknown length
 35937: contig of 1078 bp in length
 35938: gap of unknown length

36038 37158: contig of 1121 bp in length
 37159: gap of unknown length
 37259: contig of 1196 bp in length
 38455: gap of unknown length
 38455: contig of 1332 bp in length
 39787: gap of unknown length
 39886: contig of 1029 bp in length
 40915: gap of unknown length
 41015: contig of 1103 bp in length
 42119: gap of unknown length
 42218: contig of 1025 bp in length
 43244: gap of unknown length
 43344: contig of 1390 bp in length
 44733: gap of unknown length
 44834: contig of 1914 bp in length
 46747: gap of unknown length
 46848: contig of 2038 bp in length
 48885: gap of unknown length
 48886: contig of 1118 bp in length
 48986: gap of unknown length
 50104: contig of 1043 bp in length
 50204: gap of unknown length
 51247: contig of 1129 bp in length
 51347: gap of unknown length
 52476: contig of 1671 bp in length
 52576: gap of unknown length
 54247: contig of 1354 bp in length
 54347: gap of unknown length
 55701: contig of 1540 bp in length
 55801: gap of unknown length
 57341: contig of 1426 bp in length
 57441: gap of unknown length
 58667: contig of 1296 bp in length
 58967: gap of unknown length
 60263: contig of 1684 bp in length
 60362: gap of unknown length
 62046: contig of 1705 bp in length
 62147: gap of unknown length
 63852: contig of 1296 bp in length
 63951: gap of unknown length
 65247: contig of 1492 bp in length
 65348: gap of unknown length
 65348: contig of 1143 bp in length
 66840: gap of unknown length
 68083: contig of 1336 bp in length
 68182: gap of unknown length
 69518: contig of 1431 bp in length
 69519: gap of unknown length
 71049: contig of 1301 bp in length
 71050: gap of unknown length
 71150: contig of 1301 bp in length
 72450: contig of 1301 bp in length

Query Match 16.0%; Score 57.2; DB 2; Length 173786;
 Best Local Similarity 62.2%; Pred. No. 0.4;
 Matches 89; Conservative 0; Mismatches 54; Indels 0; Gaps 0;

QY 1 AAAAAAAAAAGAGTATTTTATCTCTTTTGTGATAATTTAAACA 60
 Db 27835 AAAAAAAAAAAAAAAAAATATATTTTGTGAAAAAATAATT 27894
 QY 61 TTTTGGCCCATGATATTAATAATTGATAATATATGATATGCTTT 120
 Db 27895 ATTTTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 27954
 QY 121 TTACGTCGGTGTGAGAAAGCG 143
 Db 27955 TTAAATANGGGGGGAAAGCG 27977

RESULT 7
 AC098557 309233 bp DNA linear HTG 12-JUL-2002
 LOCUS
 DEFINITION Rattus norvegicus clone CH230-81p10, *** SEQUENCING IN PROGRESS
 ACCESSION AC098557

```

VERSION      AC098557.6  GI:21728854
KEYWORDS     HTGS, HTGS_PHASE1.
SOURCE       Norway rat.
ORGANISM     Rattus norvegicus
              Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
              Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;
              Rattus.
REFERENCE    1 (bases 1 to 309233)
AUTHORS      Muzny,D.M., Adams,C., Adio-Oduola,B., Ali-osman,F.R., Allen,C.,
              Alsbrooks,S.L., Amaralung,H.C., Are,J.R., Ayele,M., Banks,T.,
              Barbarella,J., Benton,J., Blimage,K., Blankenburg,K., Bonnin,D.,
              Bouck,J., Bowie,S., Brieva,M., Brown,E., Brown,M., Bryant,N.P.,
              Buhay,C., Burch,P., Burkett,C., Burrell,K.L., Byrd,N.C.,
              Carton,T.F., Carter,M., Cavazos,S.R., Chacko,J., Chavez,D.,
              Chen,G., Chen,R., Chen,Z., Chowdhry,I., Christopoulos,C.,
              Cleveland,C.D., Cox,C., Coyle,M.D., Dathorne,S.R., David,R.,
              Davila,M.L., Davis,C., Davy-Carroll,L., Dederich,D.A.,
              Delaney,K.R., Delgado,O., Deen,A.L., Ding,Y., Dinh,H.H.,
              Douthwaite,K.J., Draper,H., Dugan-Rocha,S., Durbin,K.J.,
              Earhart,C., Edgar,D., Edwards,C.C., Elhaj,C., Escotto,M.,
              Falls,T., Ferraguto,D., Flagg,N., Ford,J., Foster,P., Frantz,P.,
              Gabisi,A., Gao,J., Garcia,A., Garner,T., Garza,N., Gill,R.,
              Gorrell,J.H., Guevara,W., Gunaratne,P., Hale,S., Hamilton,K.,
              Harris,C., Harris,K., Hart,M., Havlak,P., Hawes,A., Hernandez,J.,
              Hernandez,O., Hodgson,A., Hogues,M., Holloway,C., Hollins,B.,
              Homs,F., Howard,S., Huber,J., Hulik,S., Hume,J., Jackson,L.E.,
              Jacobson,B., Jia,Y., Johnson,R., Jolivet,S., Joudah,S.,
              Karlsson,E., Kelly,S., Khan,U., King,L., Korvah,J., Kovar,C.,
              Kratochvic,J., Kureshi,A., Landry,N., Leal,B., Lewis,L.C., Lewis,L.,
              Li,J., Li,Z., Lichtarge,O., Lieu,C., Liu,J., Liu,W., Louisedge,H.,
              Lozano,R.J., Lu,X., Lucier,A., Lucier,R., Luna,R., Ma,J.,
              Maheshwari,M., Mapua,P., Martin,R., Martindale,A., Martinez,E.,
              Massey,E., Mawhinney,E., McLeod,M.P., Meador,M., Mel,G., Metzker,M.,
              Miner,G., Neal,D., Newton,J., Newton,S., Morgan,M., Morris,S.,
              Nguyen,M., Nickerson,E., Nwokenwu,S., Oguh,M., Okunnu,G.,
              Orangunye,N., Oviedo,R., Pace,A., Payton,B., Peety,U., Perez,L.,
              Peters,L., Pichens,R., Primus,E., Pu,J.L., Quiles,M., Ren,Y.,
              Rives,M., Rojas,A., Rojudoan,I., Rolfe,M., Ruiz,S., Savery,G.,
              Scheraga,S., Scott,G., Shen,H., Shoshitari,N., Sisson,I.,
              Sodergren,E., Sonaike,T., Sparks,A., Stanley,H., Stone,H.,
              Sutton,A., Svatek,A., Taber,P., Tamerisa,A., Tamerisa,K., Tang,H.,
              Tansey,J., Taylor,C., Taylor,T., Telford,B., Thomas,R., Thomas,S.,
              Usmali,K., Vasquez,L., Vera,V., Villalobos,D., Vinson,R., Wang,Q.,
              Wang,S., Ward-Moore,S., Warren,R., Washington,C., Wallington,S.,
              Williams,G., Williamson,A., Wleczek,R., Wooden,S., Worley,K.,
              Wu,C., Wu,Y., Wu,Y.F., Zhou,J., Zorrilla,S., Nelson,D.,
              Weinstein,G. and Gibbs,R.
              Unpublished
              Direct Submission
              2 (bases 1 to 309233)
              Direct Submission
              Submitted (24-Oct-2001) Human Genome Sequencing Center, Department
              of Molecular and Human Genetics, Baylor College of Medicine, One
              Baylor Plaza, Houston, TX 77030, USA
              3 (bases 1 to 309233)
              Worley,K.C.
              Direct Submission
              Submitted (12-Jul-2002) Human Genome Sequencing Center, Department
              of Molecular and Human Genetics, Baylor College of Medicine, One
              Baylor Plaza, Houston, TX 77030, USA
              On Jul 11, 2002 this sequence version replaced gi:20066030.
              ----- Genome Center
              Center: Baylor College of Medicine
              Center code: BCM
              Web site: http://www.hgsc.bcm.tmc.edu/
              Contact: hgsc-help@bcm.tmc.edu
              ----- Project Information
              Center project name: CH230-81P10
              ----- Summary Statistics
              Sequencing vector: Plasmid;

```

Chemistry: Dye-terminator Big Dye, 100% of reads
 Assembly program: Phrap, version 0.990329
 Consensus quality: 233229 bases at least Q40
 Consensus quality: 233927 bases at least Q30
 Consensus quality: 234451 bases at least Q20

NOTE: Estimated insert size may differ from sequence length
 (see http://www.hgsc.bcm.tmc.edu/docs/Genbank_draft_data.html).
 NOTE: This is a "working draft" sequence. It currently
 consists of 61 contigs. The true order of the pieces
 is not known and their order in this sequence record is
 arbitrary. Gaps between the contigs are represented as
 runs of N, but the exact sizes of the gaps are unknown.
 This record will be updated with the finished sequence
 as soon as it is available and the accession number will
 be preserved.

```

1 1069: contig of 1069 bp in length
1070 1169: gap of unknown length
1170 2523: contig of 1354 bp in length
2524 2623: gap of unknown length
2624 3832: contig of 1209 bp in length
3833 3932: gap of unknown length
3933 4955: contig of 1023 bp in length
4956 5055: gap of unknown length
5056 6248: contig of 1193 bp in length
6249 6349: gap of unknown length
6349 7523: contig of 1175 bp in length
7524 8765: gap of unknown length
8766 10089: contig of 1142 bp in length
10090 10889: gap of unknown length
10890 11339: contig of 1224 bp in length
11340 11439: gap of unknown length
11440 12630: contig of 1150 bp in length
12631 12730: gap of unknown length
12731 13871: contig of 1191 bp in length
13872 13971: gap of unknown length
13972 14978: contig of 1141 bp in length
14978 15077: gap of unknown length
15078 16258: contig of 1006 bp in length
16259 16358: gap of unknown length
16359 17421: contig of 1181 bp in length
17422 17521: gap of unknown length
17522 18539: contig of 1063 bp in length
18540 18639: gap of unknown length
18640 19856: contig of 1018 bp in length
19857 19956: gap of unknown length
19957 20981: contig of 1217 bp in length
20982 21081: gap of unknown length
21082 22354: contig of 1025 bp in length
22355 22454: gap of unknown length
22455 23840: contig of 1273 bp in length
23841 23940: gap of unknown length
23941 25442: contig of 1366 bp in length
25443 25542: gap of unknown length
25543 26794: contig of 1502 bp in length
26795 26895: gap of unknown length
26896 28026: contig of 1252 bp in length
28027 28126: gap of unknown length
28127 29239: contig of 1132 bp in length
29240 30567: gap of unknown length
30568 31810: contig of 1113 bp in length
31811 31910: gap of unknown length
31911 33252: contig of 1442 bp in length
33253 33352: gap of unknown length
33353 34543: contig of 1191 bp in length
34544 34643: gap of unknown length
34644 35743: contig of 1000 bp in length
35744 36743: gap of unknown length
36743: contig of 1000 bp in length

```

```
* 36744 36843: gap of unknown length
* 36844 38114: contig of 1271 bp in length
* 38115 38214: gap of unknown length
* 38215 39335: contig of 1121 bp in length
* 39336 39435: gap of unknown length
* 40648 40748: contig of 1213 bp in length
* 40749 42073: gap of unknown length
* 42074 42173: gap of unknown length
* 42174 43290: contig of 1117 bp in length
* 43291 43390: gap of unknown length
* 43391 44446: contig of 1056 bp in length
* 44447 44545: gap of unknown length
* 44546 45742: contig of 1196 bp in length
* 45743 45842: gap of unknown length
* 45843 46961: contig of 1119 bp in length
* 46962 47061: gap of unknown length
* 47062 48199: contig of 1138 bp in length
* 48200 48299: gap of unknown length
* 48300 49421: contig of 1122 bp in length
* 49422 49521: gap of unknown length
* 49522 50769: contig of 1248 bp in length
* 50770 50869: gap of unknown length
* 50870 52645: contig of 1776 bp in length
* 52646 52745: gap of unknown length
* 52746 54068: contig of 1323 bp in length
* 54069 54168: gap of unknown length
* 54169 55823: contig of 1655 bp in length
* 55824 55923: gap of unknown length
* 55924 56971: contig of 1048 bp in length
* 56972 57071: gap of unknown length
* 57072 58564: contig of 1493 bp in length
* 58565 58664: gap of unknown length
* 58665 60294: contig of 1630 bp in length
* 60295 60394: gap of unknown length
* 60395 61440: contig of 1046 bp in length
* 61441 61540: gap of unknown length
* 61541 63038: contig of 1498 bp in length
* 63039 63138: gap of unknown length
* 63139 64816: contig of 1678 bp in length
* 64817 64916: gap of unknown length
* 64917 66388: contig of 1472 bp in length
* 66389 66488: gap of unknown length
* 66489 67608: contig of 1120 bp in length
* 67609 67708: gap of unknown length
* 67709 68749: contig of 1041 bp in length
* 68750 70135: gap of unknown length
* 70136 70235: contig of 1286 bp in length
* 70236 76486: contig of 6251 bp in length
```

Query Match 15.8%; Score 56.6; DB 2; Length 309233;
Best Local Similarity 63.2%; Pred. No. 0.44;
Matches 86; Conservative 0; Mismatches 50; Indels 0; Gaps 0;

```
QY 1 AAAAAAAAAAGATGAAATATTTATCTCTCTTTTGTGTAATTTAAATCA 60
DB 45162 AAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAA 45221
QY 61 TTTTTCGCCCATGATATATAAATTTGATATAATATATATGATTCGTTT 120
DB 45222 TTTTTCGTGAAAAAATAAAAAAAAAAAAAAAAAAAAAAAAAA 45281
QY 121 TTAGTTCGGGTTGAG 136
DB 45282 TTTTTCGTGTTTGGG 45297
```

RESULT 8
AX346554
LOCUS AX346554
DEFINITION Sequence 1625 from Patent WO0200928.
ACCESSION AX346554
5926 bp DNA linear PAT 01-FEB-2002

```
VERSION AX346554.1 GI:18494440  
KEYWORDS  
SOURCE synthetic construct.  
ORGANISM synthetic construct  
artificial sequences.  
REFERENCE 1  
AUTHORS Olek.A., Piepenbrock.C. and Berlin.K.  
TITLE Diagnosis of diseases associated with the immune system  
JOURNAL Patent: WO 0200928-A 1625 03-JAN-2002;  
Epigenomics AG (DE)  
FEATURES  
source location/Qualifiers  
1..5926  
/organism="synthetic construct"  
/db_xref="taxon:32630"  
/note="chemically treated genomic DNA (Homo sapiens)"  
BASE COUNT 1562 a 55 c 1318 g 2991 t  
ORIGIN
```

```
Query Match 15.8%; Score 56.4; DB 6; Length 5926;  
Best Local Similarity 48.2%; Pred. No. 1.2;  
Matches 159; Conservative 0; Mismatches 171; Indels 0; Gaps 0;  
QY 4 AAAAAAAAAAGATGATTTTATCTCTCTTTTGTGTAATTTAAATCAATT 63  
DB 3171 AAGTACAGGTCGAGTGAATTTGTTTATTTTGTGTTGTAATTTATAGAG 3230  
QY 64 TTTTTCGCCCATGATATATAAATTTGATATAATATATGATTCGTTTATA 123  
DB 3231 TTTGAATATGATTTATATATATAGTTTGTGTTGGATTTTGAATATAATTTAT 3290  
QY 124 GTTCGGGTTGAGAAAGGTTTGACCTTGGAAAGTGACGATATATGATTGGAG 183  
DB 3291 TATGAGATATTAAGTTTATGTTTGAAGAGATTTTGAATGGAATATTTGGGCTG 3350  
QY 184 CTAGTTCGAGCTTTCGACATTTGATTTGATTTGTTGTTATAGTCGACACATTT 243  
DB 3351 GGAATTTAGATATTTATTTATTTTGTTCGTTGATTTGATTAATTTAGTTTAT 3410  
QY 244 AACCCTTAATGAGCTTCTTAAGGCCCAATTAATATGATTAATCAAGTCAAC 303  
DB 3411 TTAACGAAGATTTGGTGTGTTGAGGATTTGGCGTGAAGTATAGTAT 3470  
QY 304 TTTTACCTTCGTTTGTGATCCGACCAATTA 333  
DB 3471 TATATTTGATTTTGAATGATTAATTA 3500
```

RESULT 9
AX251264
LOCUS AX251264
DEFINITION Sequence 232 from Patent WO0168912.
ACCESSION AX251264
VERSION AX251264.1 GI:15984687
KEYWORDS
SOURCE
ORGANISM synthetic construct.
synthetic construct
artificial sequences.
REFERENCE 1 (bases 1 to 12356)
AUTHORS Olek.A., Piepenbrock.C. and Berlin.K.
TITLE Diagnosis of diseases associated with tumor suppressor genes and
JOURNAL Patent: WO 0168912-A 232 20-SEP-2001;
Epigenomics AG (DE)
FEATURES
source location/Qualifiers
1..12356
/organism="synthetic construct"
/db_xref="taxon:32630"
/note="chemically treated genomic DNA (Homo sapiens)"
BASE COUNT 3645 a 68 c 2127 g 6516 t
ORIGIN

Query Match 15.5%; Score 55.6; DB 6; Length 12356;
Best Local Similarity 53.2%; Pred. No. 1.4;

and the University Cologne Institute for Biochemistry I
(http://www.uni-koeln.de/dictyostellum/project.shtml)

Funding Agency : Deutsche Forschungsgemeinschaft (DFG).
* NOTE: This is a 'working draft' sequence.
* This sequence will be replaced
* by the finished sequence as soon as it is available and
* the accession number will be preserved.

FEATURES
source . location/Qualifiers
1..93491
/organism="Dictyostelium discoideum"
/strain="AX4"
/db_xref="taxon:44689"
/chromosome="2"
/map="5401525-5495014"

BASE COUNT 36436 a 9904 c 10019 g 37032 t 100 others

ORIGIN

Query Match 15.3% Score 54.8; DB 2; Length 93491;
Best Local Similarity 50.4%; Pred. No. 1.2;
Matches 134; Conservative 0; Mismatches 132; Indels 0; Gaps 0;

OY 1 AAAAAAAAAAGATGAAGATTATTTTATTCCTCTTTTTTTTGATAATTAANCA 60
DB 65656 AAAAAAATAAAACCAATCAATCCCTTAAGATTACATTTCTTTTAAATTCCTTTT 65599
OY 61 TTTTGTGGCCAATGATATATAAAAATTTGCATATAATATTTGATATTCGTTT 120
DB 65596 TCCTTCATCATATATATTTTATTTTATTTTATTTTATATATTTTATTTTATTTT 65533
OY 121 TTACTTCGGCTTGACAAAAGCGTTTGACTTTGCAAAGTGGACGATATATGATTGC 180
DB 65536 TTTCATCATATATATTTTAAATTTGATTCGACAGGCTAATATTTTATTTTATCCTTT 6547
OY 181 GAGCTAGCTGAGCTTGGACATTCGATTTGATGTGTTGATATTTAGTGCACACT 240
DB 65476 TTTATTTTATTTTATTTTAAATTTTAAATTTTATTTTATTTTAAAAATA 6541
OY 241 ATTAACCTTAATGCGCTTTCATA 266
DB 65416 AATAAATATTTATTTGCTTTTTTTA 65391

RESULT 14
AX344570 349980 bp DNA linear PAT 01-FEB-2002
LOCUS Sequence 21 from Patent WO0200932.
DEFINITION AX344570
VERSION AX344570.1 GI:18492456
KEYWORDS synthetic construct.
SOURCE synthetic construct.
ORGANISM artificial sequences.
REFERENCE 1
AUTHORS Olek.A., Piepenbrock,C. and Berlin,K.
TITLE Diagnosis of known genetic parameters within the mhc
JOURNAL Patent: WO 0200932-A 21 03-JAN-2002;
EpiGenomics AG (DE)
location/Qualifiers
1..349980
/organism="synthetic construct"
/db_xref="taxon:32630"
/note="chemically treated genomic DNA (Homo sapiens)-Original length of seq 1: 3,673778 <22>-split as follows.-seq 01 0.000.001 TO 0.349.980-seq 02 0.300.001 649.980-seq 03 600.001 949.980-seq 04 900.001 1.249.980-seq 05 1.200.001 1.549.980-seq 06 1.500.001 1.849.980-seq 07 1.800.001 2.149.980-seq 08 2.100.001 2.449.980-seq 09 2.400.001 2.749.980-seq 10 2.700.001 3.049.980-seq 11 3.000.001 3.349.980-seq 12 3.300.001 3.649.980-seq 13 3.600.001 3.673.778 <22>-Original length of seq 2: 3,673778 <22>-split as follows.-seq 14 0.000.001 TO 0.349.980-seq 15 0.300.001 649.980-seq 16

Tumour suppressor
Human DNA for string
Tumour suppressor
Human immune system
Drosophila melanogaster
Human immune system
Human gene regulation
Tumour suppressor
Human immune system
Human immune system
Human immune system
Human immune system
Chemically pretreated
Tumour suppressor
Human immune system
DNA transcription
Tumour suppressor
Human immune system
Human immune system
Human immune system
Human immune system
Human immune system
Chemically treated
Human immune system
Human polynucleotide
Human chemically F
Human immune system
Human immune system
Human immune system
Chemically treated
Human gene regulation

```

FT exon 3914..4066
FT /*tag- h
FT Intron 4067..4164
FT /*tag- 1
FT exon 4165..4479
FT /*tag- 1
FT Intron 4480..4631
FT /*tag- k
FT exon 4632..4724
FT /*tag- 1
FT Intron 4725..4815
FT /*tag- m
FT exon 4816..4894
FT /*tag- n
FT Intron 4895..5000
FT /*tag- o
FT exon 5001..5110
FT /*tag- p
FT Intron 5111..5864
FT /*tag- q
FT exon 5865..6110
FT /*tag- r
FT 3'UTR 6011..6468
FT /*tag- s
XX
XX WO200047715-A2.
XX
XX 17-AUG-2000.
XX
XX 11-FEB-2000; 2000WO-US03820.
XX
XX 11-FEB-1999; 99US-0119657.
XX 11-FEB-1999; 99US-0119658.
XX
XX (ARIZ-) ARIZONA BOARD OF REGENTS.
XX
XX PI Azpiroz R, Choe S, Feldmann KA;
XX WPI: 2000-549142/50.
XX DR P-PSDB; AAB07921.
XX
XX PT New isolated dwf4 polynucleotide useful for altering the phenotype of
XX plants, for diagnostic assays and in the production of antibodies -
XX
XX PS Claim 3; Fig 10A-G; 113pp; English.
XX
XX CC The present sequence encodes a DWF4 polypeptide. The polypeptide is a
XX cytochrome P450 enzyme that mediates multiple steps in synthesis of
XX brassinosteroids. Specifically, it mediates multiple
XX 22alpha-hydroxylation steps in brassinosteroid biosynthesis. The DWF4
XX polynucleotide is used for altering the phenotype of a plant. DWF4
XX plants display a dramatic reduction in the length of different organs,
XX and this size reduction is attributable to a defect in cell elongation.
XX The DWF4 polynucleotides and polypeptides can be used in diagnostic
XX assays and to generate antibodies, which can be used to produce
XX immunogenic compositions.
XX
XX SQ Sequence 6888 BP; 2294 A; 1010 C; 1193 G; 2391 T; 0 other;
XX
XX Query Match 100.0%; Score 358; DB 21; Length 6888;
XX Best Local Similarity 100.0%; Pred. No. 3.2e-55;
XX Matches 358; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
XX
XX QY 1 AAAAAAAAAAGATGAAGTATTTTATCTCTCTTTTGTGATTAATTTAAATCA 60
XX |||||||
XX Db 6111 AAAAAAAAAAGATGAAGTATTTTATCTCTCTTTTGTGATTAATTTAAATCA 6170
XX
XX QY 61 TTTTTCGCCCAATGATATATAATAATTTGGATAAATATATTGATTCGTTTT 120
XX |||||||
XX Db 6171 TTTTTCGCCCAATGATATATAATAATTTGGATAAATATATTGATTCGTTTT 6230
XX
XX QY 121 TTGTCGGGTTGAGAAAGCGTTTGACTTCGAAAGCGAGCATGTATAGATTGG 180
XX |||||||

```

```

Db 6231 TTAGTTCGGGTTTGAGAAAAGGTTTCGACTTTCGAAAGTGAGCATGTATAGATTGG 6290
QY 181 GAGCTAGGTTGAGCTTTTGACATTTGATTTGATTTGATTTAGTTCGACACT 240
|||
Db 6291 GAGCTAGGTTGAGCTTTTGACATTTGATTTGATTTGATTTAGTTCGACACT 6350
QY 241 ATTTAAACCTTAAGTGGCTTTCTATTAAGGCCCAATTAATTAATTAACAAAGTGAC 300
|||
Db 6351 ATTTAAACCTTAAGTGGCTTTCTATTAAGGCCCAATTAATTAATTAACAAAGTGAC 6410
QY 301 AACTTTTACTGCTTTTGAGCCGAGCAATTAACAAATTCGAAATTAACAAACAGC 358
|||
Db 6411 AACTTTTACTGCTTTTGAGCCGAGCAATTAACAAATTCGAAATTAACAAACAGC 6468

RESULT 2
ABL33641
ID ABL33641 standard; DNA; 5449 BP.
XX
XX ABL33641;
XX
XX AC 26-MAR-2002 (first entry)
XX
XX DE Human immune system associated gene SEQ ID NO: 1614.
XX
XX KW Human; immune system disease; cytosine methylation; antiasthmatic;
XX antiarteriosclerotic; antianaemic; cytosolic; nocrotic;
XX neuroprotective; anti-HIV; anticonvulsant; ophthalmological;
XX antirheumatic; antiarthritic; antidiabetic; antipsoriatic;
XX antinflammatory; cancer; eye disease; arteriosclerosis; anaemia;
XX acute myeloid leukaemia; Alzheimer's disease; AIDS; epilepsy;
XX neurofibromatosis; rheumatoid arthritis; psoriasis; bowel disease;
XX gene; ds.
XX
XX OS Homo sapiens.
XX
XX FN WO200200928-A2.
XX
XX PD 03-JAN-2002.
XX
XX PF 02-JUL-2001; 2001WO-EP07537.
XX
XX PR 30-JUN-2000; 2000DE-1032529.
XX PR 01-SEP-2000; 2000DE-1043826.
XX
XX PA (EPIG-) EPIGENOMICS AG.
XX
XX PI Olek A, Plepenbrock C, Berlin K;
XX
XX DR WPI: 2002-130909/17.
XX
XX PT Nucleic acid comprising fragment of chemically modified gene, useful
XX for diagnosis and treatment of diseases associated with abnormal
XX cytosine methylation -
XX
XX PS Claim 1; SEQ ID NO 1614; 32pp + Sequence Listing; German.
XX
XX CC The present invention provides a number of human immune system associated
XX genes which are modified by the methylation of cytosines. The sequences
XX can be used in the diagnosis and treatment of immune system disorders,
XX including eye diseases such as retinopathy, neovascular glaucoma and
XX macular degeneration, arteriosclerosis, anaemia, cancer, acute myeloid
XX leukaemia, Alzheimer's disease, AIDS, epilepsy, neurofibromatosis,
XX rheumatoid arthritis, psoriasis and inflammatory/ulcerative bowel
XX diseases. The present sequence is a gene of the invention.
XX
XX SQ Sequence 5449 BP; 1146 A; 90 C; 1299 G; 2914 T; 0 other;
XX
XX Query Match 16.2%; Score 58; DB 24; Length 5449;
XX Best Local Similarity 65.4%; Pred. No. 0.053;
XX Matches 85; Conservative 0; Mismatches 45; Indels 0; Gaps 0;
XX
XX QY 4 AAAAAAAAAAGATGAAGTATTTTATCTCTCTTTTGTGATTAATTTAAATCAATT 63

```

DB 5256 AAAATATAAGTATTTTGAATTTTATTTTATTTTATTTTAACTT 5315
OY 64 TTTTGGCCATGATATATATAATTTGATTAATATTTGATTCGTTTAA 123
DB 5316 TGTATAAGTTTATATATATTTTATATATTTTAAATTTGATTTTATTT 5375
OY 124 GTTCGGCTT 133
DB 5376 TTTTGGCTT 5385

RESULT 3
ABL33223
ID ABL33223 standard; DNA; 7560 BP.
XX
AC ABL33223;
XX
DT 26-MAR-2002 (first entry)
XX
DE Human immune system associated gene SEQ ID NO: 1196.
XX
KW Human; immune system disease; cytosine methylation; antiasthmatic;
KW antiarteriosclerotic; antianaemic; cytosolic; noctropic;
KW neuroprotective; anti-HIV; anticonvulsant; ophthalmological;
KW antirheumatic; antiarthritic; antidiabetic; antipsoriatic;
KW antineoplastic; cancer; eye disease; arteriosclerosis; anaemia;
KW acute myeloid leukaemia; Alzheimer's disease; AIDS; epilepsy;
KW neurofibromatosis; rheumatoid arthritis; psoriasis; bowel disease;
KW gene; ds.
XX
OS Homo sapiens.
XX
PN WO200200928-A2.
XX
PD 03-JAN-2002.
XX
PE 02-JUL-2001; 2001WO-EP07537.
XX
PR 30-JUN-2000; 2000DE-1032529.
PR 01-SEP-2000; 2000DE-1043826.
XX
PA (EPIG-) EPIGENOMICS AG.
XX
PI Olek A. Piepenbrock C, Berlin K.
XX
DR WPI: 2002-130909/17.
XX
PT Nucleic acid comprising fragment of chemically modified gene, useful
PT for diagnosis and treatment of diseases associated with abnormal
PT cytosine methylation
XX
PS Claim 1; SEQ ID NO 1196; 32pp + Sequence Listing; German.
XX
CC The present invention provides a number of human immune system associated
CC genes which are modified by the methylation of cytosines. The sequences
CC can be used in the diagnosis and treatment of immune system disorders,
CC including eye diseases such as retinopathy, neovascular glaucoma and
CC macular degeneration, arteriosclerosis, anaemia, cancer, acute myeloid
CC leukaemia, Alzheimer's disease, AIDS, epilepsy, neurofibromatosis,
CC rheumatoid arthritis, psoriasis and inflammatory/ulcerative bowel
CC diseases. The present sequence is a gene of the invention.
XX
SQ Sequence 7560 BP; 2154 A; 47 C; 1554 G; 3805 T; 0 other;

Query Match 16.2%; Score 58; DB 24; Length 7560;
Best Local Similarity 51.6%; Pred. No. 0.053;
Matches 159; Conservative 0; Mismatches 145; Indels 4; Gaps 1;

OY 2 AAAAAAAAAAGTATTTTATCTCTCTTTTATTTTATTTTAAATCAT 61
DB 2728 AATAATTAGATAAATTTTATTTTATTTTATTTTATTTAGATTATTTA 2787

OY 62 TTTTGGCCATGATATATATAATTTGATTAATATTTGATTCGTTTAA 121
DB 2788 TTTTATGTGAGATTTTAAATTTTAAAGATAATTTTATTTTATGCTGTAATT 2847
OY 122 TAGTTGGGTTTGAGAAAAGGTTTCGACTTCGAAAGTGACATGATTTAGATTGGG 181
DB 2848 TGTATTTGATTTAAATTAATTTTATTTAGATTTTATTTGATTTTAAATAAATTAAT 2907
OY 182 AGCTAGCTGTAGCTCTTTCGACATTTGATTTGAGTGTGATTTAGTGTGAC 237
DB 2908 TATTTATTAATTAATTAATTAATTAATTAATTAATTAATTAATTTTAAATTT 2967
OY 238 ACTATTAACCTTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 297
DB 2968 ATATTTGATTTAAATTAATTTTATTTTAAAGAAATTTATGGAATTAATTAAGA 3027
OY 298 GACAACTT 305
DB 3028 GATAAATT 3035

RESULT 4
ABL33652
ID ABL33652 standard; DNA; 5926 BP.
XX
AC ABL33652;
XX
DT 26-MAR-2002 (first entry)
XX
DE Human immune system associated gene SEQ ID NO: 1625.
XX
KW Human; immune system disease; cytosine methylation; antiasthmatic;
KW antiarteriosclerotic; antianaemic; cytosolic; noctropic;
KW neuroprotective; anti-HIV; anticonvulsant; ophthalmological;
KW antirheumatic; antiarthritic; antidiabetic; antipsoriatic;
KW antineoplastic; cancer; eye disease; arteriosclerosis; anaemia;
KW acute myeloid leukaemia; Alzheimer's disease; AIDS; epilepsy;
KW neurofibromatosis; rheumatoid arthritis; psoriasis; bowel disease;
KW gene; ds.
XX
OS Homo sapiens.
XX
PN WO200200928-A2.
XX
PD 03-JAN-2002.
XX
PE 02-JUL-2001; 2001WO-EP07537.
XX
PR 30-JUN-2000; 2000DE-1032529.
PR 01-SEP-2000; 2000DE-1043826.
XX
PA (EPIG-) EPIGENOMICS AG.
XX
PI Olek A. Piepenbrock C, Berlin K.
XX
DR WPI: 2002-130909/17.
XX
PT Nucleic acid comprising fragment of chemically modified gene, useful
PT for diagnosis and treatment of diseases associated with abnormal
PT cytosine methylation
XX
PS Claim 1; SEQ ID NO 1625; 32pp + Sequence Listing; German.
XX
CC The present invention provides a number of human immune system associated
CC genes which are modified by the methylation of cytosines. The sequences
CC can be used in the diagnosis and treatment of immune system disorders,
CC including eye diseases such as retinopathy, neovascular glaucoma and
CC macular degeneration, arteriosclerosis, anaemia, cancer, acute myeloid
CC leukaemia, Alzheimer's disease, AIDS, epilepsy, neurofibromatosis,
CC rheumatoid arthritis, psoriasis and inflammatory/ulcerative bowel
CC diseases. The present sequence is a gene of the invention.
XX
SQ Sequence 5926 BP; 1562 A; 55 C; 1318 G; 2991 T; 0 other;

Query Match 15.8%; Score 56.4; DB 24; Length 5926;
 Best Local Similarity 48.2%; Pred. No. 0.1;
 Matches 159; Conservative 0; Mismatches 171; Indels 0; Gaps 0;

4 AAAAAAGATGAAGATTTTATCTCTTTTGTGATTTTAAATCATT 63
 3171 AAGATGAGGTTCGATTTGTTTATTTTATTTTGAATTTTATAGAG 3230
 64 TTTTGGCCATGATATATAAATTTGATTAATATATGATTTCTTTTAA 123
 3231 TTTGAATGTTATTTATATAGTTTGTGTTGGATTTTGAATTAATTTAT 3250
 124 GTTCGGTTGAGAAAAGGCTTCGACTTCGAAAGTGACGATGATATAGTTGGAG 183
 3291 TATGAGATATTAAGTTTACGTTTGAAGAAGATTTGGAATGAAATTTGGGTG 3350
 184 CTAGGTGAGCTTTTGACATTTGATTTGATTTGATTTATAGTCGACATATT 243
 3351 GGAGTTTATATATTTATTTATTTTGTGATTTGATTTGATTTTATTTAT 3410
 244 AATCTTAAATGGCTTTCTATTAAGGCCAATTAATTAATTAACAAAGTACAC 303
 3411 TTAACGAAGTGGGTGGTGGAGATGGGTAGGATGAGGTTGGAAGATATAGTAT 3470
 304 TTTTACTCTCTTTTGTATCCGAAGCATTA 333
 3471 TATATTTGATTTTGAATGATTTTAAATTA 3500

RESULT 5

AAS46510
 ID AAS46510 standard; DNA: 12356 BP.

AAS46510:
 18-DEC-2001 (first entry)
 Tumour suppressor gene derived chemically modified sequence #232.
 Human; tumour suppressor gene; oncogene; antitumour; cytosolic;
 cancer; tumour; CpG dinucleotide; single-nucleotide polymorphism; SNP;
 cytosine methylation; ds.

OS Homo sapiens.
 XX
 PN WO200168912-A2.

PD 20-SEP-2001.

PF 15-MAR-2001; 2001WO-EP02955.

XX 15-MAR-2000; 2000DE-1013847.
 PR 06-APR-2000; 2000DE-1019058.
 PR 07-APR-2000; 2000DE-1019173.
 PR 30-JUN-2000; 2000DE-1032529.
 PR 01-SEP-2000; 2000DE-1043826.

XX (EPiG-) EPIGENOMICS AG.

PI Olek A, Piepenbrock C, Berlin K;

DR WPI; 2001-602752/68.

XX Fragments of chemically modified genes associated with tumour suppressor
 PT genes and oncogenes, useful in designing primers and probes for
 PT analysing diseases associated with cytosine methylation state e.g.
 cancer

XX Claim 1; SEQ ID No 232; 27pp; English.

CC The invention relates to a nucleic acid comprising a sequence of 18
 CC bases, of a segment of chemically pretreated DNA (CP DNA) e.g. with

CC bisulphite, of genes associated with tumour suppression and
 CC oncogenes having a sequence taken from 536 (actually 533 since
 CC numbers 408, 458 and 500 are missing from the sequence listing) sequences
 CC (Ss) and sequences complementary to (Ss). The nucleic acid may be a
 CC peptide nucleic acid-oligomer (PNA) of at least 9 nucleotides and may
 CC form part of a set of probes for detecting the cytosine methylation state
 CC and/or single nucleotide polymorphisms and also to be used in an
 CC array for analysing diseases associated with CpG dinucleotides e.g.
 CC cancers and tumours. The probes can also be used in a method for
 CC ascertaining genetic and/or epigenetic parameters for the diagnosis
 CC and/or therapy of existing diseases or the predisposition to specific
 CC diseases, by analysing cytosine methylations. The parameters may be
 CC compared to another set of genetic and/or epigenetic parameters, the
 CC differences serving as basis for diagnosis and/or prognosis events which
 CC are disadvantageous to patients. The present sequence is one of the
 CC 533 genomic sequences derived from tumour suppressor genes and
 CC oncogenes. Sequences with even numbered Seq ID numbers are the
 CC complementary sequence of the corresponding odd numbered sequence (e.g.
 CC ID 2 and ID1, ID 536 and ID 535, except for those whose partner sequence
 CC is missing).

CC Note: The sequence data for this patent did not form part
 CC of the printed specification, but was obtained in electronic
 CC format directly from WIPO at
 CC ftp.wipo.int/pub/published_pct_sequences.

XX Sequence 12356 BP; 3645 A; 68 C; 2127 G; 6516 T; 0 other;

Query Match 15.5%; Score 55.6; DB 22; Length 12356;
 Best Local Similarity 53.2%; Pred. No. 0.14;
 Matches 118; Conservative 0; Mismatches 104; Indels 0; Gaps 0;

QY 11 AGATGAAGATTTTATCTCTCTTTTGTGATTAATTTAAATCATTCTTTTGC 70
 9051 AGATTAATGCTAGTTATTTTATTTTGAAGATTTTAAATGTTTATATAGT 9110
 QY 71 CCAATGATATTAATAATTTGATTAATAATATATATGATTCGTTTATAGTGGG 130
 9111 GGTGTATGATTAATTTTATTTTAAATATATAGGCTTTTATTTTATATGTT 9170
 QY 131 TTTGCAAAAAGGCTTTCGACTTTCGAAAGTGACAGATATATAGTGGAGCTAGGTT 190
 9171 TATTAGATTTTGTATATAGTTTGTGATTAATAGTATTTAAATGGTAGATGAT 9230
 QY 191 GAGTCTTGGACATTTGATTTGATGATGTTGATTTAGTG 232
 9231 ATTTTATTTGATTTTGTATTTTGTATTTTGTGATGATTAATG 9272

RESULT 6

ABL32028
 ID ABL32028 standard; DNA: 7351 BP.

XX ABL32028;

DE 26-MAR-2002 (first entry)

XX Human immune system associated gene SEQ ID NO: 1.

XX Human; immune system disease; cytosine methylation; antiasthmatic;
 KW antiarteriosclerotic; antianemic; cytosolic; neurotropic;
 KW neuroprotective; anti-HIV; anticonvulsant; ophthalmological;
 KW antineumatic; antiarthritic; antidiabetic; antiporotic;
 KW antineumatic; antiarthritic; antidiabetic; antiporotic;
 KW acute myeloid leukaemia; cancer; eye disease; arteriosclerosis; anaemia;
 KW neurofibromatosis; rheumatoid arthritis; psoriasis; bowel disease;
 KW gene; ds.

XX Homo sapiens.

XX WO200200928-A2.

XX 03-JAN-2002.

KW acute myeloid leukemia; Alzheimer's disease; AIDS; epilepsy;
 KW neurofibromatosis; rheumatoid arthritis; psoriasis; bowel disease;
 KW gene; ds.
 XX Homo sapiens.
 XX WO200200928-A2.
 PN 03-JAN-2002.
 PD 03-JAN-2002.
 XX 02-JUL-2001; 2001WO-EP07537.
 PF 30-JUN-2000; 2000DE-1032529.
 PR 01-SEP-2000; 2000DE-1043826.
 XX (EPIC-) EPIGENOMICS AG.
 PA Olek A, Piepenbrock C, Berlin K;
 PI WPI; 2002-130909/17.
 DR Nucleic acid comprising fragment of chemically modified gene, useful
 XX for diagnosis and treatment of diseases associated with abnormal
 PT cytosine methylation
 PS Claim 1; SEQ ID NO 2100; 32pp + Sequence Listing; German.
 XX The present invention provides a number of human immune system associated
 CC genes which are modified by the methylation of cytosines. The sequences
 CC can be used in the diagnosis and treatment of immune system disorders,
 CC including eye diseases such as retinopathy, neovascular glaucoma and
 CC macular degeneration, arteriosclerosis, anemia, cancer, acute myeloid
 CC leukemia, Alzheimer's disease, AIDS, epilepsy, neurofibromatosis,
 CC rheumatoid arthritis, psoriasis and inflammatory/ulcerative bowel
 CC diseases. The present sequence is a gene of the invention.
 XX
 SO Sequence 2814 BP; 629 A; 144 C; 770 G; 1271 T; 0 other;
 Query Match 15.1%; Score 54.2; DB 24; Length 2814;
 Best Local Similarity 48.8%; Pred. No. 0.26;
 Matches 146; Conservative 0; Mismatches 153; Indels 0; Gaps 0;
 QY 20 TATTTTATCTCTCTTTTGGATTAATTTAAATCTTTTGGCCCAATATA 79
 DB 3 TTTGTTTAAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 62
 QY 80 TATTAATAATTTGGATTAATTAATTAATTTGTTTAAATTTTGGAGAAA 139
 DB 63 TAGGTGAGGTGATATATTTTGTATATATTTGTTATATGTAATTTTGTAT 122
 QY 140 AGCGTTTCAGCTTCGAAAGTGCAGATGATATAGATTGGAGCTAGTGTG 199
 DB 123 ATCTATATGTTATATATATATGTTATGTTATGTTTGTATTAACGTTTATA 182
 QY 200 GACATTTGTTGATTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTT 259
 DB 183 TGTATATATATTAATTAATTAATTTTATTAATTAATTAATTAATTAATTA 242
 QY 260 TTCTATAGAGCCCAATTAATTAATTAATTAATTAATTAATTAATTAATTA 318
 DB 243 TTTGGAGAGTTTAAATTAATTAATTAATTAATTAATTAATTAATTAATTA 301
 RESULT 9
 ABR28455
 ID ABR28455 standard; DNA; 7138 BP.
 XX ABR28455;
 AC
 XX 23-APR-2002 (first entry)
 DT
 XX DNA transcription associated genomic DNA #165.
 DE
 XX

KW DNA transcription associated gene; peptide nucleic acid; PNA-oligomer;
 KW PNA; cytosine methylation state; SNP; retroviral infection; gene; ds;
 KW single nucleotide polymorphism; adenosine deaminase deficiency; cancer;
 KW viral infection; Sezary syndrome; haematological disorder; tuberculosis;
 KW immunological disorder; Werner syndrome; developmental disorder;
 KW psoriasis; Rieger's syndrome; neurological disorder; erythropoiesis;
 KW neurodegenerative disorder; Mardenburg syndrome; Niemann-Pick disease;
 KW myelodysplastic syndrome; myocardial infarction; hypertension; arthritis;
 KW angiodysplasia; congenital heart disease; HDR syndrome; gene therapy;
 KW polyglutamine disorder; solid tumour.
 XX
 OS Unidentified.
 XX
 XX WO200192565-A2.
 PN 06-DEC-2001.
 PD 06-DEC-2001.
 XX 06-APR-2001; 2001WO-EP03973.
 PF 06-APR-2000; 2000DE-1019058.
 PR 07-APR-2000; 2000DE-1019173.
 PR 30-JUN-2000; 2000DE-1032529.
 PR 01-SEP-2000; 2000DE-1043826.
 XX (EPIC-) EPIGENOMICS AG.
 PA Olek A, Piepenbrock C, Berlin K;
 PI WPI; 2002-090046/12.
 DR New nucleic acids or oligomers, useful for diagnosing or treating
 XX diseases associated with DNA transcription, e.g. immunological
 PT disorders, Werner syndrome, psoriasis, myocardial infarction, solid
 PT tumours or cancer
 PS Claim 1; SEQ ID NO 329; 32pp; English.
 XX The invention relates to a nucleic acid, which comprises a segment of the
 CC chemically pretreated DNA of genes associated with DNA transcription from
 CC one of 346 sequences, and an oligomer, in particular an oligonucleotide
 CC or peptide nucleic acid (PNA)-oligomer that hybridises to or is identical
 CC to the chemically pretreated DNA of genes associated with DNA
 CC transcription. The set of oligomer probes are useful for detecting the
 CC cytosine methylation state and/or single nucleotide polymorphisms (SNPs)
 CC in a chemically pretreated genomic DNA. The nucleic acids are useful for
 CC diagnosing or treating diseases associated with DNA transcription
 CC (particularly with the methylation status), e.g. adenosine deaminase
 CC deficiency, viral infection, retroviral infection, Sezary syndrome,
 CC haematological disorders, immunological disorders, Werner syndrome,
 CC tuberculosis, developmental disorders, psoriasis, Rieger's syndrome,
 CC neurological disorders, neurodegenerative disorders, Mardenburg
 CC syndrome, Niemann-Pick disease, myelodysplastic syndrome, myocardial
 CC infarction, hypertension, angiodysplasia, erythropoiesis, congenital heart
 CC disease, HDR syndrome, arthritis, polyglutamine disorders, solid tumours
 CC or cancer. Sequences ABR28127-ABR28472 represent DNA transcription
 CC associated genomic DNA molecules of the invention.
 CC Note: The sequence data for this patent did not form part of the printed
 CC specification but was obtained in electronic format directly from the
 CC European Patent Office.
 XX
 SO Sequence 7138 BP; 1564 A; 288 C; 1759 G; 3527 T; 0 other;
 Query Match 15.1%; Score 54; DB 24; Length 7138;
 Best Local Similarity 48.1%; Pred. No. 0.27;
 Matches 153; Conservative 0; Mismatches 165; Indels 0; Gaps 0;
 QY 3 AAAAAAAAAAGATGAAATATTTATCTCTCTTTTGGATTAATTTAAATCAT 62
 DB 3163 AAAAAATTAATAGAGAGGCTGTGTTATTTTATATTTTGGTTATTTGGATTGTT 3222
 QY 63 TTTTTCGCCCAATGATATATTAATAAATTTGATTAATTAATTAATTAATTAATTA 122
 DB 3223 TTTTGGATTAATGTTATTTTATTTTGTATTTTATTTTATTTGTTTGTGTTT 3282

OY 123 AGTTCGGGTTGAGAAAAGGTTTCGACTTCGAAAGTGACAGATGATATAGATTGGGA 182
 DB 3283 AATTAGTTTAAAGATGATATTTGATATATATAGATGATTTTATATAGTATGAA 3342
 OY 183 GCTAGGTTGAGCTCTTTGGACATTTGTATTTGATGTTGATTTAGTGTGACACATAT 242
 DB 3343 ATATTTTATTTTATTTTATTTTATTTGATTTTATTTAGTGTATTTTTCGGGGTAA 3402
 OY 243 TAAACCTTAATGGGCTTCTATTAAGGCCCAATTTATATAGATATATACAAAGTGACAA 302
 DB 3403 TTAAGTATATTTTATTTTATTTGAGGTTTGTGTTTGTATTTTAAATTTTATAGAT 3462
 OY 303 CTTTACTTCGTTTGA 320
 DB 3463 ATTTTATAGTTTATTA 3480
 RESULT 10
 AAS46743
 ID AAS46743 standard; DNA; 14147 BP.
 AC AAS46743;
 DT 18-DEC-2001 (first entry)
 DE Tumour suppressor gene derived chemically modified sequence #467.
 KM Human; tumour suppressor gene; oncogene; antitumour; cytostatic;
 KM cancer; tumour; CpG dinucleotide; single-nucleotide polymorphism; SNP;
 KM cytosine methylation; ds.
 OS Homo sapiens.
 PN WO200168912-A2.
 PD 20-SEP-2001.
 PF 15-MAR-2001; 2001WO-EP02955.
 PR 15-MAR-2000; 2000DE-1013847.
 PR 06-APR-2000; 2000DE-1019058.
 PR 07-APR-2000; 2000DE-1019173.
 PR 30-JUN-2000; 2000DE-1032529.
 PR 01-SEP-2000; 2000DE-1043826.
 PA (EPIC-) EPIGENOMICS AG.
 PI Olek A, Piepenbrock C, Berlin K;
 DR WPI; 2001-602752/68.
 PT Fragments of chemically modified genes associated with tumour suppressor
 PT genes and oncogenes, useful in designing primers and probes for
 PT analysing diseases associated with cytosine methylation state e.g.
 PT cancer
 PS Claim 1: SEQ ID No 467; 27pp; English.
 XX The invention relates to a nucleic acid comprising a sequence of 18
 CC bases, of a segment of chemically pretreated DNA (CP DNA) e.g. with
 CC bisulphite, of genes associated with tumour suppression and
 CC oncogenes having a sequence taken from 536 (actually 533 since
 CC numbers 408, 458 and 500 are missing from the sequence listing) sequences
 CC (Ss) and sequences complementary to (Ss). The nucleic acid may be a
 CC peptide nucleic acid-oligomer (PNA) of at least 9 nucleotides and may
 CC form part of a set of probes for detecting the cytosine methylation state
 CC and/or single nucleotide polymorphisms and also to be used in an
 CC array for analysing diseases associated with CpG dinucleotides e.g.
 CC cancers and tumours. The probes can also be used in a method for
 CC ascertaining genetic and/or epigenetic parameters for the diagnosis
 CC and/or therapy of existing diseases or the predisposition to specific
 CC diseases, by analysing cytosine methylations. The parameters may be

CC compared to another set of genetic and/or epigenetic parameters, the
 CC differences serving as basis for diagnosis and/or prognosis events which
 CC are disadvantageous to patients. The present sequence is one of the
 CC 533 genomic sequences derived from tumour suppressor genes and
 CC oncogenes.
 CC Note: The sequence data for this patent did not form part
 CC of the printed specification, but was obtained in electronic
 CC format directly from WIPO at
 CC ftp.wipo.int/pub/published_pcl_sequences.
 SQ Sequence 14147 BP; 3621 A; 225 C; 3116 G; 7185 T; 0 other;
 Query Match 14.9%; Score 53.2; DB 22; Length 14147;
 Best Local Similarity 48.1%; Pred. No. 0.36;
 Matches 151; Conservative 0; Mismatches 163; Indels 0; Gaps 0;
 OY 7 AAAAAGATGAAAGATTTTATTTCTCTCTTTTGTGATTAATTTAATCAATTTT 66
 DB 10811 AAAAAGTTTGGTTAGTTGTTTGGTTTGGTTATTTGTTTTCGTTGTTT 10870
 OY 67 TTGCCCATGATATATAAAATTTGGATATAATATATATGATATTCGTTTACTT 126
 DB 10871 TTTTTCGAATTTATAGTTAGTTTATTTTATGTTATTTGTTAGCTATTTTACTT 10930
 OY 127 CGGTTTGAGAAAGGTTTCGATTTGGAAGTGACAGATATATAGATTGGAGCTA 186
 DB 10931 TTGGGTTTATATACGTTGCGTTTATTTATTTGTTGTTTATTTAGATTTTACAG 10990
 OY 187 GGTGAGTCTTTTGACATTTGTTGATGTTGTTGATTTATAGTGTGCACACTATTTAA 246
 DB 10991 TTTAGTTTATTTATTTTATTTTGTGTTTGTGTTAGTTGTTTGTGCGATATTTT 11050
 OY 247 CCTTAATGGGCTTCTATTAAGGCCCAATTAATATATACATTTAACAAGTACACTT 306
 DB 11051 TTTTGATTTTATTAAGTAAAGTAAATAGTTTATTTGTTGATTTTATTAATTT 11110
 OY 307 TACTTCGTTTGA 320
 DB 11111 TTATTTATTTTGA 11124
 RESULT 11
 ABRK3955
 ID ABRK3955 standard; DNA; 14147 BP.
 AC ABRK3955;
 DT 18-JUN-2002 (first entry)
 DE Human DNA for staging of Astrocytomas #20.
 KM Human; ds; astrocytoma; cytostatic; staging; cysteine methylation; CpG;
 KM bisulphite; brain tissue; MALDI; ESI; electron spray mass spectrometry;
 KM matrix assisted laser desorption/ionization mass spectrometry.
 OS Homo sapiens.
 PN WO200202808-A2.
 PD 10-JAN-2002.
 PF 02-JUL-2001; 2001WO-EP07538.
 PR 30-JUN-2000; 2000DE-1032529.
 PR 01-SEP-2000; 2000DE-1043826.
 PA (EPIC-) EPIGENOMICS AG.
 PI Olek A, Piepenbrock C, Berlin K;
 DR WPI; 2002-171649/22.
 PT Novel chemically modified genomic DNA sequences, useful in the

PT Characterisation, classification, differentiation, grading, staging,
PT treatment and/or diagnosis of astrocytomas or predisposition to
PT astrocytomas -
XX
XX
PS Claim 1; SEQ ID No 39; 37pp; English.

The invention relates to a nucleic acid comprising a sequence (I) of at least 18 bases in length of a segment of chemically pre-treated genomic DNA which has any one of the sequences of (ARK31915-ABK34032) or its complement. Also included are an oligonucleotide or peptide nucleic acid (or set thereof) of at least 9 nucleotides which hybridises to (I), primers for (I), probes for detecting cytosine methylation or single-nucleotide polymorphisms (SNP) in (I), an array of oligomers or peptide nucleic acids for analysing diseases associated with the methylation states of the CpG dinucleotides of (I). The array is useful for determining genetic and/or epigenetic parameters, classification, differentiation, grading, staging, treatment and/or diagnosis of astrocytomas, or the predisposition to astrocytomas by analysing cytosine methylations, involves obtaining a biological sample containing genomic DNA, extracting the genomic DNA, converting cytosine bases which are unmethylated at the 5-position, in the genomic DNA sample, to uracil or another base which is dissimilar to cytosine in terms of hybridisation behaviour, by chemical treatment and amplifying chemically pre-treated genomic DNA fragments using the array and a polymerase, where the amplicates carry a detectable label. The method further involves identifying methylation status of one or more cytosine positions, and analysing methylation status of the cytosine positions by reference to one or more data sets. The genomic DNA is chemically treated by using a bisulphite, hydrogen sulphite or disulphite. The amplification step amplifies DNA which is of particular interest in astrocytoma or brain tissue, based on the specific genomic methylation status of brain tissues, as opposed to background DNA. The amplicates carry a fluorescent label or radionucleide. Optionally, the labels of the amplicates are detachable molecule fragments having a typical mass which are detected in a mass spectrometer. The fragments of chemically pre-treated genomic DNA to be amplified, have a single positive or negative charge for a better detectability in the mass spectrometer. Preferably, the amplicates or fragments of the amplicates are detected by matrix assisted laser desorption/ionization mass spectrometry (MALDI) or using electron spray mass spectrometry (ESI). The present sequence is one of the chemically pre-treated reference DNA samples of the invention.

Note: The sequence data for this patent did not form part of the printed specification, but was obtained in electronic format directly from WIPO at http://wipo.int/pub/published_pct_sequences.

Sequence	14147	BP;	3621	A;	225	C;	3116	G;	7185	T;	0	other;
Query Match	14.9%	Score	53.2;	DB	24;	Length	14147;					
Best Local Similarity	48.1%;	Pred. No.	0.36;									
Matches	151;	Conservative	0;	Mismatches	163;	Indels	0;	Gaps	0;			

QY	7	AAAAAGTGAAGATATTTATCTCTCTCTTTTTTTTGATTAATTAACATTTTT	66
Db	10811	AAAAAGTTGGTTAGTTGGTTTCGTTATTTTGTTTTTTTTTCGTTGTTTTTTT	10870
QY	67	TTGCCCAATATATATAAAATTTGGATAATAATATATGATATCGTTTATAGTT	126
Db	10871	TTTTTTTCGAATTTATAGTTATTTTTTTTTTTTATGTTATTTGATTTTAAAGTT	109330
QY	127	CGGGTTTGACAAAAGGTTTCGACTTCGAAAGTCAGCATATATATAGTGGAGCTA	186
Db	10931	TTGGGTTTTTAAATCGGTTGTCGTTTTTATTTTGGTTGTGTTTTTAATATTTACGAG	109900
QY	187	GGTGAAGCTTTTGGACATTTGTATTTGATGTTGTTGATTAATAGTGCACACTATTAAA	246
Db	10991	TTTAGTTTTTTTATTTTTTTTTTTTGGTTTGGTTTAGTTGTTTTTTTTTGGCGATATATTT	110500
QY	247	CCCTAAATGGCCTTTCATATAGGCCCAATATATTTACGATTAATACAAAGTGACACTTT	306
Db	11051	TTTTTGATTTTTTAATGTAAGTAATAGTTTTTTTTTGTGATTTTTTTTTTATATATATTT	111100

QY	307	TACTTCGTTTTGA	320
Db	11111	TTATTATTATTGA	11124

RESULT 12
 ID AAS46670
 XX AAS46670 standard; DNA; 7261 BP.
 AC AAS46670;
 XX
 DT 18-DEC-2001 (first entry)
 XX
 DE Tumour suppressor gene derived chemically modified sequence #392.
 XX
 KW Human; tumour suppressor gene; oncogene; antitumour; cytostatic;
 KW cancer; tumour; CpG dinucleotide; single-nucleotide polymorphism; SNP;
 KW cytosine methylation; ds.
 XX
 OS Homo sapiens.
 XX
 PN WO200168912-A2.
 XX
 PD 20-SEP-2001.
 XX
 PF 15-MAR-2001; 2001WO-EP029955.
 XX
 PR 15-MAR-2000; 2000DE-1013847.
 PR 06-APR-2000; 2000DE-1019058.
 PR 07-APR-2000; 2000DE-1019173.
 PR 30-JUN-2000; 2000DE-1033529.
 PR 01-SEP-2000; 2000DE-1043826.
 XX
 PA (EPIC-) EPIGENOMICS AG.
 XX
 PI Olek A. Plepenbrock C, Berlin K;
 XX
 DR MPI; 2001-602752/68.
 XX
 PT Fragments of chemically modified genes associated with tumour suppressor
 PT genes and oncogenes, useful in designing primers and probes for
 PT analysing diseases associated with cytosine methylation state e.g.
 XX cancer -
 PS
 PS Claim 1; SEQ ID No 392; 27pp; English.
 XX
 CC The invention relates to a nucleic acid comprising a sequence of 18
 CC bases, of a segment of chemically pretreated DNA (CP DNA) e.g. with
 CC bisulphite, of genes associated with tumour suppression and
 CC oncogenes having a sequence taken from 336 (actually 333 since
 CC numbers 408, 458 and 500 are missing from the sequence listing) sequences
 CC (Ss) and sequences complementary to (Ss). The nucleic acid may be a
 CC peptide nucleic acid-oligonmer (PNA) of at least 9 nucleotides and may
 CC form part of a set of probes for detecting the cytosine methylation state
 CC and/or single nucleotide polymorphisms and also to be used in an
 CC array for analysing diseases associated with CpG dinucleotides e.g.
 CC cancers and tumours. The probe can also be used in a method for
 CC ascertaining genetic and/or epigenetic parameters for the diagnosis
 CC and/or therapy of existing diseases or the predisposition to specific
 CC diseases, by analysing cytosine methylations. The parameters may be
 CC compared to another set of genetic and/or epigenetic parameters, the
 CC differences serving as basis for diagnosis and/or prognosis events which
 CC are disadvantageous to patients. The present sequence is one of the
 CC 533 genomic sequences derived from tumour suppressor genes and
 CC oncogenes. Sequences with even numbered Seq ID numbers are the
 CC complementary sequence of the corresponding odd numbered sequence (e.g.
 CC ID 2 and ID1, ID 536 and ID 535, except for those whose partner sequence
 CC is missing).
 CC Note: The sequence data for this patent did not form part
 CC of the printed specification, but was obtained in electronic
 CC format directly from WIPO at
 CC ftp.wipo.int/pub/published_pct_sequences.
 CC

Sequence 7261 BP; 1862 A; 171 C; 1854 G; 3374 T; 0 other;

Query Match 14.8%; Score 53; DB 22; Length 7261;

Best Local Similarity 50.6%; Pred. No. 0.41; Mismatches 125; Indels 0; Gaps 0;

Matches 128; Conservative 0; Mismatches 125; Indels 0; Gaps 0;

OY 13 ATGAAGATATTTTATCTCTCTTTTGTATTAATTTTAAATCATTTTGGCC 72
 DB 6407 ATAAAGAAATATTTTATTTTAAATGTTATTTTGTATTTTGTATTTT 6466
 OY 73 AATGATATATATAATTTGATTAATATATATTTGATTTTCTTTTGTGCGGTT 132
 DB 6467 CGTATATATATATATATTAAGAAAGAAAGATAGTATTTAGTTTGTATTT 6526
 OY 133 TGAGAAAGAGGTTTCGACTTCGAAAGAGAGATGATATAGATTTGGAGCTAGTTGA 192
 DB 6527 TGTAGAAATATTAAGAGCTGCGATTTTGTAGTGTGATTAAGTTATAGGTAAAGGA 6586
 OY 193 GTCTTTGACATTTGTATTTGATGTTTGTATTTACTGTGACACTATTTAACTTAA 252
 DB 6587 TTTTATTTTATTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGA 6646
 OY 253 ATGGGCTTCTAT 265
 DB 6647 TTTGTTTTTTTTT 6659

RESULT 13

ABL33138
 ID ABL33138 standard; DNA; 5678 BP.

AC ABL33138;

XX 26-MAR-2002 (first entry)

DE Human immune system associated gene SEQ ID NO: 1111.

XX Human: immune system disease; cytosine methylation; antiasthmatic;
 KM antileukosclerotic; antianemic; cytosine methylation; noctropic;
 KM antineoplastic; anti-HIV; anticonvulsant; ophthalmological;
 KM antirheumatic; antirheumatic; antidiabetic; arteriosclerosis; anaemia;
 KM antineoplastic; cancer; eye disease; arteriosclerosis; anaemia;
 KM acute myeloid leukaemia; Alzheimer's disease; AIDS; epilepsy;
 KM neurofibromatosis; rheumatoid arthritis; psoriasis; bowel disease;
 KM gene; ds.

OS Homo sapiens.

XX WO200200928-A2.

XX 03-JAN-2002.

XX 02-JUL-2001; 2001WO-EP07537.

XX 30-JUN-2000; 2000DE-1032529.

XX 01-SEP-2000; 2000DE-1043826.

XX (EPIC-) EPIGENOMICS AG.

XX Olek A, Piepenbrock C, Berlin K;

XX WPI; 2002-130909/17.

XX Nucleic acid comprising fragment of chemically modified gene, useful
 PT for diagnosis and treatment of diseases associated with abnormal
 PT cytosine methylation

XX Claim 1; SEQ ID NO 1111; 32bp + Sequence Listing; German.

XX The present invention provides a number of human immune system associated
 CC genes which are modified by the methylation of cytosines. The sequences
 CC can be used in the diagnosis and treatment of immune system disorders,
 CC including eye diseases such as retinopathy, neovascular glaucoma and

CC macular degeneration, arteriosclerosis, anaemia, cancer, acute myeloid
 CC leukaemia, Alzheimer's disease, AIDS, epilepsy, neurofibromatosis,
 CC rheumatoid arthritis, psoriasis and inflammatory/ulcerative bowel
 CC diseases. The present sequence is a gene of the invention.

Sequence 5678 BP; 1387 A; 32 C; 1284 G; 2975 T; 0 other;

Query Match 14.7%; Score 52.8; DB 24; Length 5678;

Best Local Similarity 54.0%; Pred. No. 0.44; Mismatches 92; Indels 0; Gaps 0;

Matches 108; Conservative 0; Mismatches 92; Indels 0; Gaps 0;

OY 36 TTTTATTTTATTTTATTTTAAATCATTTTGTCCCAATGATATATAAATTTGGATA 95
 DB 3308 TTTTATTTTATTTTATTTTAAATGTTTGTATTTTGTATTTTGTATTTTGTG 3367
 OY 96 AATATATATTTGATATCTTTTGTATTTGTTGATTAAGAAAGGTTTGCATTTG 155
 DB 3368 AATTTTATTTTATTTTATTTTAAATGTTTGTATTTGTTGTTGTTGTTGTTG 3427
 OY 156 AAGTGCAGATGATATAGATTTGGAGCTAGTTGATCTTTGACATTTGATTTGAT 215
 DB 3428 TATTTTATTTTATTTTATTTTATTTTATTTTATTTTATTTTATTTTATTTGAG 3487
 OY 216 GTTGTGATTTATTTAGTGTG 235
 DB 3488 TATGTTGATTTTATTTTGTG 3507

RESULT 14

ABL19002
 ID ABL19002 standard; DNA; 21231 BP.

AC ABL19002;

XX 26-MAR-2002 (first entry)

DE Drosophila melanogaster genomic polynucleotide SEQ ID NO 8479.

XX Drosophila: developmental biology; cell signalling; insecticide;
 KM pharmaceutical; gene; ds.

OS Drosophila melanogaster.

XX WO200171042-A2.

XX 27-SEP-2001.

XX 23-MAR-2001; 2001WO-US09231.

XX 23-MAR-2000; 2000US-191637P.

XX 11-JUL-2000; 2000US-0614150.

XX (PEKE) PE CORP NY.

XX Venter JC, Adams M, Li PWD, Myers EW;

XX WPI; 2001-656860/75.

XX New isolated nucleic acid detection reagent for detecting 1000 or more
 PT genes from Drosophila and for elucidating cell signalling and cell-cell
 PT interactions

XX Claim 1; SEQ ID NO 8479; 21bp + Sequence Listing; English.

XX The invention relates to an isolated nucleic acid detection reagent
 CC capable of detecting 1000 or more genes from Drosophila. The invention is
 CC useful in developmental biology and in elucidating cell signalling and
 CC cell-cell interactions in higher eukaryotes for the development of
 CC insecticides, therapeutics and pharmaceutical drugs. The invention
 CC discloses genomic DNA sequences (ABL16176-ABL30511), expressed DNA
 CC sequences (ABL57737-ABL72072) and the encoded proteins
 CC (ABL57737-ABL72072).
 CC The sequence data for this patent did not form part of the printed

CC specification, but was obtained in electronic format directly from WIPO
 CC at ftp.wipo.int/pub/published_pct_sequences.

XX Sequence 21231 BP; 6727 A; 3859 C; 3768 G; 6877 T; 0 other;

Query Match 14.7%; Score 52.6; DB 23; Length 21231;
 Best Local Similarity 46.8%; Pred. No. 0.46;
 Matches 166; Conservative 0; Mismatches 189; Indels 0; Gaps 0;

QY 1 AAAAAAAAAAAGATGAAGTATTTTATCTCTCTTTTGTGATTAATTTAAATCA 60
 DB 16810 AAAAGTCAGACAGCAATTAATAAACTGCTGATTTTATTTATTTATTTAT 16869
 QY 61 TTTTTCGCCCAATGATATATATAAAATTTGGATTAATTAATTTGATATTCGTTT 120
 DB 16870 ATATATATAGAAATATATATATATATTTTAAACAATGATCTATATATACAGTTT 16929
 QY 121 TTACTGGGTTTCAGAAAAGGTTTCGACTTTCGAAAGTGACGATATATGATTGG 180
 DB 16930 GAAAAAATTTTAAAGTATTTATTTATTTGGAAGATGATGCTTAAATATTTGT 16989
 QY 181 GAGCTAGCTGAGCTTTGGACATTTGATTTGATGTTGATTATTTAGTCGACACT 240
 DB 16990 AATTAATTTTTCACCTTTGTAATATAGATCTGATTTAATTAAGTCATTAATTCG 17049
 QY 241 ATTAACCTTAATGAGCTTTCTATTAAGGCCCAATTAATTAATTAACAAGTGAC 300
 DB 17050 GCCAGACCAAAAGTCCGTTTTCGAAATCGAAAAATTTTACGATTTTGAACGAG 17109
 QY 301 AACTTTACTGCTTTTGTGATCCGACCAATTAACAATTCGAATACCAACAC 355
 DB 17110 CATCGATTTTCTATTTCAATGGAATTTGAAATCTTACGATTCGATTAAC 17164

RESULT 15

ABL32322
 ID ABL32322 standard; DNA: 6418 BP.

XX ABL32322;

XX 26-MAR-2002 (first entry)

DE Human immune system associated gene SEQ ID NO: 295.

XX Human; immune system disease; cytosine methylation; antiasthmatic;
 KW antiarteriosclerotic; antianemic; cytosolic; noctropic;
 KW neuroprotective; anti-HIV; anticonvulsant; ophthalmological;
 KW antineumatic; antiarthritic; antidiabetic; antipsoriatic;
 KW antinflammatory; cancer; eye disease; arteriosclerosis; anaemia;
 KW acute myeloid leukaemia; Alzheimer's disease; AIDS; epilepsy;
 KW neurofibromatosis; rheumatoid arthritis; psoriasis; bowel disease;
 KW gene; ds.

XX Homo sapiens.

XX W0200200928-A2.

XX 03-JAN-2002.

XX 02-JUL-2001; 2001WO-EP07537.

XX 30-JUN-2000; 2000DE-1032529.

XX 01-SEP-2000; 2000DE-1043826.

XX (EPig-) EPIGENOMICS AG.

XX Olek A, Piepenbrock C, Berlin K;

XX WPI; 2002-130909/17.

XX Nucleic acid comprising fragment of chemically modified gene, useful
 PT for diagnosis and treatment of diseases associated with abnormal
 PT cytosine methylation

XX Claim 1; SEQ ID NO 295; 32pp + Sequence Listing; German.

XX The present invention provides a number of human immune system associated
 CC genes which are modified by the methylation of cytosines. The sequences
 CC can be used in the diagnosis and treatment of immune system disorders,
 CC including eye diseases such as retinopathy, neovascular glaucoma and
 CC macular degeneration, arteriosclerosis, anaemia, cancer, acute myeloid
 CC leukaemia, Alzheimer's disease, AIDS, epilepsy, neurofibromatosis,
 CC rheumatoid arthritis, psoriasis and inflammatory/ulcerative bowel
 CC diseases. The present sequence is a gene of the invention.

XX Sequence 6418 BP; 1842 A; 47 C; 1423 G; 3106 T; 0 other;

Query Match 14.6%; Score 52.2; DB 24; Length 6418;
 Best Local Similarity 56.8%; Pred. No. 0.57;
 Matches 96; Conservative 0; Mismatches 73; Indels 0; Gaps 0;

QY 18 AGATTTTATCTCTCTTTTGTGATTAATTTTAAATCAATTTTGGCCCAATGA 77
 DB 1110 AATTTTATTTAGTTTGTGTTTATTTTATTTATTTATTTTGTGTTAGTTG 1169
 QY 78 TATATAAAATTTGATTAATTAATTTTGTGATTTTCGTTTATTTAGTTGGGTTGAGA 137
 DB 1170 TATTTATATTTGTAGTTTAATTAATTAATTAATTTTGTGTTTGTAGATAAGTA 1229
 QY 138 AAGGCTTCGACTTTCGAAAGTGACGATGATTAATGATTGGAGCTA 186
 DB 1230 TAAATGATTTATTTTGTAGCGGATTAATTTATATGCTTTTATTTA 1278

Search completed: June 10, 2003, 22:22:49
 Job time : 83.9537 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: June 10, 2003, 22:02:58 ; Search time 17.0513 Seconds
(without alignments)
6438.831 Million cell updates/sec

Title: US-09-502-426B-1_COPY_6111_6468

Perfect score: 358
Sequence: 1 aaaaaaaaaagatgaaagt.....tgcaaatcaacaacacaag 358

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 441362 seqs, 153338381 residues

Total number of hits satisfying chosen parameters: 882724

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database :

Issued_Patents_NA:*
1: /cgn2_6/ptodata/1/lna/5A.COMB.seq:*
2: /cgn2_6/ptodata/1/lna/5B.COMB.seq:*
3: /cgn2_6/ptodata/1/lna/6A.COMB.seq:*
4: /cgn2_6/ptodata/1/lna/6B.COMB.seq:*
5: /cgn2_6/ptodata/1/lna/PCITS.COMB.seq:*
6: /cgn2_6/ptodata/1/lna/Dackfilest1.seq:*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	length	ID	Description
1	42.8	12.0	1493	US-08-340-820-24	Sequence 24, Appl
2	42.8	12.0	1493	US-08-593-535-24	Sequence 24, Appl
3	41.8	11.7	998	US-09-122-400B-5	Sequence 5, Appl
4	41	11.5	2230	US-08-378-313-24	Sequence 24, Appl
5	40.8	11.4	6243	US-09-056-075-1	Sequence 1, Appl
6	40.8	11.4	19557	PCR-US92-06300-1	Sequence 1, Appl
7	40.6	11.3	1733	US-09-073-569-1	Sequence 1, Appl
8	40.6	11.3	3095	5231168-1	Patent No. 5231168
9	40.4	11.3	10607	US-08-078-090-3	Sequence 3, Appl
10	40	11.2	2755	US-08-749-522-2	Sequence 2, Appl
11	39.4	11.0	289	US-08-341-568-3	Sequence 3, Appl
12	39.4	11.0	289	US-08-911-020-3	Sequence 3, Appl
13	39.4	11.0	19124	US-08-487-826B-13	Sequence 13, Appl
14	39	10.9	1895	US-09-444-336-7	Sequence 7, Appl
15	39	10.9	4285	US-09-410-464-1	Sequence 1, Appl
16	39	10.9	9048	US-08-973-273-4	Sequence 4, Appl
17	38.8	10.8	835	US-08-998-416-547	Sequence 547, App
18	38.8	10.8	863	US-08-998-416-498	Sequence 498, Appl
19	38.8	10.8	3138	US-07-867-106-4	Sequence 4, Appl
20	38.2	10.7	2836	US-08-747-221B-24	Sequence 24, Appl
21	38.2	10.7	2836	US-08-747-221B-26	Sequence 26, Appl
22	38.2	10.7	2836	US-09-005-051-24	Sequence 24, Appl
23	38.2	10.7	2836	US-09-005-051-26	Sequence 26, Appl
24	38.2	10.7	5852	US-07-867-106-2	Sequence 2, Appl
25	38.2	10.7	29604	US-08-781-891-207	Sequence 207, App
26	38	10.6	552	US-09-134-001C-2606	Sequence 2606, Ap
27	38	10.6	701	US-08-998-416-701	Sequence 701, App

28	38	10.6	2555	2	US-08-693-457-3	Sequence 3, Appl
29	38	10.6	2555	4	US-09-265-731-3	Sequence 3, Appl
30	38	10.6	8700	2	US-08-392-625-16	Sequence 16, Appl
31	38	10.6	8700	2	US-08-466-961A-16	Sequence 16, Appl
32	38	10.6	8700	2	US-08-645-193B-18	Sequence 18, Appl
33	37.8	10.6	1582	3	US-08-545-196B-10	Sequence 10, Appl
34	37.8	10.6	1582	3	US-08-545-196B-12	Sequence 12, Appl
35	37.6	10.5	2107	4	US-09-180-852-1	Sequence 1, Appl
36	37.6	10.5	2422	1	US-07-867-106-5	Sequence 5, Appl
37	37.6	10.5	3138	1	US-07-867-106-4	Sequence 4, Appl
38	37.6	10.5	5852	1	US-07-867-106-2	Sequence 2, Appl
39	37.6	10.5	6078	4	US-09-173-914-1	Sequence 1, Appl
40	37.4	10.4	477	4	US-08-887-534A-81	Sequence 81, Appl
41	37.4	10.4	2445	6	5215909-9	Patent No. 5215909
42	37.4	10.4	2674	4	US-09-817-180-1	Sequence 1, Appl
43	37.4	10.4	3761	4	US-08-890-865A-2	Sequence 2, Appl
44	37.4	10.4	10223	4	US-08-961-527-73	Sequence 73, Appl
45	37.2	10.4	144	1	US-08-702-344-26	Sequence 26, Appl

ALIGNMENTS

RESULT 1
US-08-340-820-24/C
Sequence 24, Application US/08340820
Patent No. 5512460
GENERAL INFORMATION:
APPLICANT: NARUO, Ken-ichi
APPLICANT: SEKO, Chisako
APPLICANT: KUROKAWA, Tsutomu
APPLICANT: KONDO, Tatsuya
TITLE OF INVENTION: GLIA ACTIVATING FACTOR AND ITS
NUMBER OF SEQUENCES: 27
CORRESPONDENCE ADDRESS:
ADDRESSEE: DAVID G. CONLIN; DIKE, BRONSTEIN, ROBERTS &
ADDRESS: CUSHMAN
STREET: 130 Water Street
City: Boston
STATE: Massachusetts
COUNTRY: US
ZIP: 02109
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/340,820
FILING DATE:
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US/07/835,713
FILING DATE:
ATTORNEY/AGENT INFORMATION:
NAME: CONLIN, David G.
REGISTRATION NUMBER: 27026
TELECOMMUNICATION INFORMATION:
TELEPHONE: (617)523-3400
TELEFAX: (617)523-6440
TELEX: 200291 STRE UR
INFORMATION FOR SEQ ID NO: 24:
SEQUENCE CHARACTERISTICS:
LENGTH: 1493 base pairs
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: linear
MOLECULE TYPE: CDNA to mRNA
HYPOTHETICAL: NO
ANTI-SENSE: NO
ORIGINAL SOURCE:
ORGANISM: Homo sapiens

? HAPLOTYPE: 2n
 ? TISSUE TYPE: skin
 ? CELL TYPE: fibroblast
 ? IMMEDIATE SOURCE:
 ? LIBRARY: Human foreskin CDNA library
 ? CLONING VECTOR: pGAT1
 ?
 US-08-340-820-24

Query Match	12.0%;	Score 42.8;	DB 1;	Length 1493;
Best Local Similarity	57.5%;	Pred. No. 0.94;		
Matches 77;	Conservative	0;	Mismatches 57;	Indels 0;
				Gaps 0

QY	20	TATTTATATCTCTCTTTTATTTATATTTAATCAATTTTGGCCAAATGTA	79
Db	1469	TTTGAAT	1410
QY	80	TATATAATTTGCATAATAATATATTTGGATATTCGTTTTAGTGGGTTGAGAA	139
Db	1409	TCTATAATAATAATTAACTTTTTTATTTTATTTTATTTTATTTTATTTTACIA	1350
QY	140	AGGCTTCGACTTT	153
Db	1349	GAGGCTTTTGCTTT	1336

RESULT 2
US-08-593-535-24/C

1 GENERAL INFORMATION:
2 APPLICANT: NARUO, Ken-ichi
3 APPLICANT: SEKO, Chisako
4 APPLICANT: KUROKAWA, Tsutomu
5 APPLICANT: KONDO, Tatsuya
6 TITLE OF INVENTION: GLIA ACTIVATING FACTOR AND ITS
7 TITLE OF INVENTION: PRODUCTION
8 NUMBER OF SEQUENCES: 27
9 CORRESPONDENCE ADDRESS:

1
 2
 3
 4
 5
 6
 7
 8
 9
 10
 11
 12
 13
 14
 15
 16
 17
 18
 19
 20
 21
 22
 23
 24
 25
 26
 27
 28
 29
 30
 31
 32
 33
 34
 35
 36
 37
 38
 39
 40
 41
 42
 43
 44
 45
 46
 47
 48
 49
 50
 51
 52
 53
 54
 55
 56
 57
 58
 59
 60
 61
 62
 63
 64
 65
 66
 67
 68
 69
 70
 71
 72
 73
 74
 75
 76
 77
 78
 79
 80
 81
 82
 83
 84
 85
 86
 87
 88
 89
 90
 91
 92
 93
 94
 95
 96
 97
 98
 99
 100
 101
 102
 103
 104
 105
 106
 107
 108
 109
 110
 111
 112
 113
 114
 115
 116
 117
 118
 119
 120
 121
 122
 123
 124
 125
 126
 127
 128
 129
 130
 131
 132
 133
 134
 135
 136
 137
 138
 139
 140
 141
 142
 143
 144
 145
 146
 147
 148
 149
 150
 151
 152
 153
 154
 155
 156
 157
 158
 159
 160
 161
 162
 163
 164
 165
 166
 167
 168
 169
 170
 171
 172
 173
 174
 175
 176
 177
 178
 179
 180
 181
 182
 183
 184
 185
 186
 187
 188
 189
 190
 191
 192
 193
 194
 195
 196
 197
 198
 199
 200
 201
 202
 203
 204
 205
 206
 207
 208
 209
 210
 211
 212
 213
 214
 215
 216
 217
 218
 219
 220
 221
 222
 223
 224
 225
 226
 227
 228
 229
 230
 231
 232
 233
 234
 235
 236
 237
 238
 239
 240
 241
 242
 243
 244
 245
 246
 247
 248
 249
 250
 251
 252
 253
 254
 255
 256
 257
 258
 259
 260
 261
 262
 263
 264
 265
 266
 267
 268
 269
 270
 271
 272
 273
 274
 275
 276
 277
 278
 279
 280
 281
 282
 283
 284
 285
 286
 287
 288
 289
 290
 291
 292
 293
 294
 295
 296
 297
 298
 299
 300
 301
 302
 303
 304
 305
 306
 307
 308
 309
 310
 311
 312
 313
 314
 315
 316
 317
 318
 319
 320
 321
 322
 323
 324
 325
 326
 327
 328
 329
 330
 331
 332
 333
 334
 335
 336
 337
 338
 339
 340
 341
 342
 343
 344
 345
 346
 347
 348
 349
 350
 351
 352
 353
 354
 355
 356
 357
 358
 359
 360
 361
 362
 363
 364
 365
 366
 367
 368
 369
 370
 371
 372
 373
 374
 375
 376
 377
 378
 379
 380
 381
 382
 383
 384
 385
 386
 387
 388
 389
 390
 391
 392
 393
 394
 395
 396
 397
 398
 399
 400
 401
 402
 403
 404
 405
 406
 407
 408
 409
 410
 411
 412
 413
 414
 415
 416
 417
 418
 419
 420
 421
 422
 423
 424
 425
 426
 427
 428
 429
 430
 431
 432
 433
 434
 435
 436
 437
 438
 439
 440
 441
 442
 443
 444
 445
 446
 447
 448
 449
 450
 451
 452
 453
 454
 455
 456
 457
 458
 459
 460
 461
 462
 463
 464
 465
 466
 467
 468
 469
 470
 471
 472
 473
 474
 475
 476
 477
 478
 479
 480
 481
 482
 483
 484
 485
 486
 487
 488
 489
 490
 491
 492
 493
 494
 495
 496
 497
 498
 499
 500
 501
 502
 503
 504
 505
 506
 507
 508
 509
 510
 511
 512
 513
 514
 515
 516
 517
 518
 519
 520
 521
 522
 523
 524
 525

1
 2
 3
 4
 5
 6
 7
 8
 9
 10
 11
 12
 13
 14
 15
 16
 17
 18
 19
 20
 21
 22
 23
 24
 25
 26
 27
 28
 29
 30
 31
 32
 33
 34
 35
 36
 37
 38
 39
 40
 41
 42
 43
 44
 45
 46
 47
 48
 49
 50
 51
 52
 53
 54
 55
 56
 57
 58
 59
 60
 61
 62
 63
 64
 65
 66
 67
 68
 69
 70
 71
 72
 73
 74
 75
 76
 77
 78
 79
 80
 81
 82
 83
 84
 85
 86
 87
 88
 89
 90
 91
 92
 93
 94
 95
 96
 97
 98
 99
 100
 101
 102
 103
 104
 105
 106
 107
 108
 109
 110
 111
 112
 113
 114
 115
 116
 117
 118
 119
 120
 121
 122
 123
 124
 125
 126
 127
 128
 129
 130
 131
 132
 133
 134
 135
 136
 137
 138
 139
 140
 141
 142
 143
 144
 145
 146
 147
 148
 149
 150
 151
 152
 153
 154
 155
 156
 157
 158
 159
 160
 161
 162
 163
 164
 165
 166
 167
 168
 169
 170
 171
 172
 173
 174
 175
 176
 177
 178
 179
 180
 181
 182
 183
 184
 185
 186
 187
 188
 189
 190
 191
 192
 193
 194
 195
 196
 197
 198
 199
 200
 201
 202
 203
 204
 205
 206
 207
 208
 209
 210
 211
 212
 213
 214
 215
 216
 217
 218
 219
 220
 221
 222
 223
 224
 225
 226
 227
 228
 229
 230
 231
 232
 233
 234
 235
 236
 237
 238
 239
 240
 241
 242
 243
 244
 245
 246
 247
 248
 249
 250
 251
 252
 253
 254
 255
 256
 257
 258
 259
 260
 261
 262
 263
 264
 265
 266
 267
 268
 269
 270
 271
 272
 273
 274
 275
 276
 277
 278
 279
 280
 281
 282
 283
 284
 285
 286
 287
 288
 289
 290
 291
 292
 293
 294
 295
 296
 297
 298
 299
 300
 301
 302
 303
 304
 305
 306
 307
 308
 309
 310
 311
 312
 313
 314
 315
 316
 317
 318
 319
 320
 321
 322
 323
 324
 325
 326
 327
 328
 329
 330
 331
 332
 333
 334
 335
 336
 337
 338
 339
 340
 341
 342
 343
 344
 345
 346
 347
 348
 349
 350
 351
 352
 353
 354
 355
 356
 357
 358
 359
 360
 361
 362
 363
 364
 365
 366
 367
 368
 369
 370
 371
 372
 373
 374
 375
 376
 377
 378
 379
 380
 381
 382
 383
 384
 385
 386
 387
 388
 389
 390
 391
 392
 393
 394
 395
 396
 397
 398
 399
 400
 401
 402
 403
 404
 405
 406
 407
 408
 409
 410
 411
 412
 413
 414
 415
 416
 417
 418
 419
 420
 421
 422
 423
 424
 425
 426
 427
 428
 429
 430
 431
 432
 433
 434
 435
 436
 437
 438
 439
 440
 441
 442
 443
 444
 445
 446
 447
 448
 449
 450
 451
 452
 453
 454
 455
 456
 457
 458
 459
 460
 461
 462
 463
 464
 465
 466
 467
 468
 469
 470
 471
 472
 473
 474
 475
 476
 477
 478
 479
 480
 481
 482
 483
 484
 485
 486
 487
 488
 489
 490
 491
 492
 493
 494
 495
 496
 497
 498
 499
 500
 501
 502
 503
 504
 505
 506
 507
 508
 509
 510
 511
 512
 513
 514
 515
 516
 517
 518
 519
 520
 521
 522
 523
 524
 525

? BILLING DATE: 12-FEB-1992
 ? ATTORNEY/AGENT INFORMATION:
 ? NAME: CONLIN, David G.
 ? REGISTRATION NUMBER: 27026
 ? TELECOMMUNICATION INFORMATION:
 ? TELEPHONE: (617)523-3400
 ? TELEFAX: (617)523-6440
 ? TELETYPE: 200291 STRE NO
 ? INFORMATION FOR SEQ ID NO: 24:
 ? SEQUENCE CHARACTERISTICS:
 ? LENGTH: 1493 base pairs

```

; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
; MOLECULE TYPE: cDNA to mRNA
; HYPOTHETICAL: NO
; ANTI-SENSE: NO
;

```

? ORIGINAL SOURCE:
 ? ORGANISM: Homo sapiens
 ? HAPLOTYPE: 2n
 ? TISSUE TYPE: skin
 ? CELL TYPE: fibroblast
 ? IMMEDIATE SOURCE:
 ? LIBRARY: Human foreskin CDNA library
 ? CLONE: PGAT1
 ? US-08-593-335-24

Query Match	12.0%; Score 42.8; DB 1; Length 1493;
-------------	---------------------------------------

Matches	77;	Conservative	0;	Mismatches	57;	Indels	0;	Gaps	0;
QY	20	TATTTTATTCCTCTCTTTT	TTTTTTTGATATTTTAATCA	TTTTTTTGGCCCAATGATA	79				
Db	1469	TTTTTTTTTTTTTTTTTTTT	TTTTTTTTTTTTTTTTTTTT	TTTTTTTTTTTTTTTGGAAAT	1410				
QY	80	TATAAATTTGATAATAATA	TATTATTTGATATTCGTTT	TATAGTCGGTTGAGANA	139				
Db	1409	TCATTAATTAATTAATTA	ACTTTTTTAAATTTTATTT	TATTTTATTTTAACTTAA	1350				
QY	140	AGCGTTTCGACTTT	153						
Db	1349	GAGCTCTTGCTT	1336						

RESULT 3

; Sequence 5, Application US/09122400E

```

1  GENERATED INFORMATION:
2  APPLICANT: Michalowski, Susan
3  APPLICANT: Spitzer, Steven
4  TITLE OF INVENTION: MATRIX ATTACHMENT REGIONS
5  FILE REFERENCE: Michalowski and Spitzer
6  CURRENT APPLICATION NUMBER: US/09/122,400B
7  CURRENT FILING DATE: 1998-07-24
8  PRIOR APPLICATION NUMBER: 60/066,118
9  PRIOR FILING DATE: 1997-08-06
10 NUMBER OF SEQ ID NOS: 22
11 SOFTWARE: Patentin Ver. 2.1

```

```

; SEQ ID NO 5
; LENGTH: 998
; TYPE: DNA
; ORGANISM: Nicotiana tabacum
US-09-122-400B-5

```

Query Match	11.7%	Score 41.8;	DB 4;	Length 998;
Best Local Similarity	50.2%;	Pred. No. 1.5;		
Matches 103;	Conservative 0;	Mismatches 102;	Indels 0;	Gaps 0;

QY	5	AAAAAATGATGAAAGATATTTTATTTATCTCTCTTTTATTTTGTAAATTTAAATCAATTT	64
Db	619	AAAAAAGACAGAAATATTTTTTTTTTGTAGTTTTTACAAAATATGTCTTTAGAAAATAT	56CC
QY	65	TTTGGCCAAATGATATATATAAAATTTGCAATAATATATTATGTATATTCGTTTTTAG	124A
Db	559	TTCAAGCTTTTTTTTAAAGCAGTTTTTTTGCTAAATAATAGAAAAAAAATATTTTCATTTT	500C
QY	125	TTTCGGGTTTGGAAAAAGGTTTCGACTTTTCGAAGTGCACGATGTATATAGATTGGGAGC	184A
Db	499	TTTCAGTTTTTTTGTAGTTTTTTTCCAGATTTTACAAAATATAATTCCTTAGAAAATTCATTT	440C
QY	185	TAGGTGAGTCTTTGGACATTTGTA	209
Db	439	TCAGCTCTCTTTTTTTTCAGTTTTTA	415

RESULT 4
US-08-378-313-24/C
; Sequence 24, Application US/08378313
; Patent No. 6207881

RESULT 4
US-08-378-313-24/C
; Sequence 24, Application US/08378313
; Patent No. 6207881

GENERAL INFORMATION:

APPLICANT: THEOLOGIS, ATHANASIOS
 APPLICANT: SATO, TAKAHITO
 TITLE OF INVENTION: CONTROL OF FRUIT RIPENING THROUGH
 TITLE OF INVENTION: GENETIC CONTROL OF ACC SYNTHASE SYNTHESIS
 NUMBER OF SEQUENCES: 34
 CORRESPONDENCE ADDRESS:
 ADDRESSEE: MORRISON & FOERSTER
 STREET: 755 Page Mill Road
 CITY: Palo Alto
 STATE: California
 COUNTRY: USA
 ZIP: 94304-1018
 COMPUTER READABLE FORM:
 MEDIUM TYPE: Floppy disk
 COMPUTER: IBM PC compatible
 OPERATING SYSTEM: PC-DOS/MS-DOS
 SOFTWARE: Patentin Release #1.0, Version #1.25
 CURRENT APPLICATION DATA:
 APPLICATION NUMBER: US/08/378,313
 FILING DATE:
 CLASSIFICATION: 800
 PRIOR APPLICATION DATA:
 APPLICATION NUMBER: US 07/862,493
 FILING DATE: 02-APR-1992
 ATTORNEY/AGENT INFORMATION:
 NAME: MURASHIGE, KATE H.
 REGISTRATION NUMBER: 29,959
 REFERENCE/DOCKET NUMBER: 29190-20002.20
 TELECOMMUNICATION INFORMATION:
 TELEPHONE: (415) 856-5600
 TELEFAX: (415) 494-0792
 TELEX: 706141
 INFORMATION FOR SEQ ID NO: 24:
 SEQUENCE CHARACTERISTICS:
 LENGTH: 2230 base pairs
 TYPE: nucleic acid
 STRANDEDNESS: single
 TOPOLOGY: linear
 FEATURE:
 NAME/KEY: CDS
 LOCATION: 91..1545
 US-08-378-313-24

Query Match 11.5%; Score 41; DB 4; Length 2230;
 Best Local Similarity 44.9%; Pred. No. 2.1; Indels 0; Gaps 0;

Matches 155; Conservative 0; Mismatches 190; Indels 0; Gaps 0;

4 AAAAAAAGATGAAGATATTTTATTCCTCTTTTGGATTAATTAATCATTT 63
 2222 AACAATATATTTTATTTTATTTTATTTTATTTTATTTTATTTTATTTT 2163
 64 TTTTGGCCCAATGATATATAAATTTGGATTAATTAATTAATTAATTTT 123
 2162 TTTTATTAATAGCTAAGCAAAATTAATGATCTAAGTATTTTATTTT 2103
 124 GTTCGGCTTGAGAAAGGCTTCGACTTCGAAGTGAGCATGATTAATGCGAG 183
 2102 TTAACATTAATCTGACTTTGCAATCAATTAATAGCTAAGCAAAATTAATGATCTAAGTA 2043
 184 CTAGTTGAGCTTTGGACATTTGATGATGATGATGATGATGATGATGATGAT 243
 2042 CATTTTATTTTATTTTATTTTATTTTATTTTATTTTATTTTATTTTATTT 1983
 244 AAACTTAATGGGCTTCTTAATAGGCCAATTAATTAATGATTAATTAACAAAGTGACAAAC 303
 1982 ACAAAATTAATGATGCTAAGTACATTTTATTTTATTTTATTTTATTTTATTTTATTTTATTT 1923
 304 TTTTACTGCTTTTGAATGCGAAGCAATTAACAAATTAATGATTAATTAACAAAGTGACAAAC 348
 1922 TTTGATTCATTAATTAATGCTAAGCAAAATTAATGATGCTAAGTAC 1878

RESULT 5

US-09-056-075-1/c

Sequence 1, Application US/09056075

Patent No. 5953368

GENERAL INFORMATION:

APPLICANT: Johnson, Eric A.
 APPLICANT: Bradshaw, Marile
 APPLICANT: Reed, Julian
 TITLE OF INVENTION: Expression System for Clostridium
 TITLE OF INVENTION: Species
 NUMBER OF SEQUENCES: 2
 CORRESPONDENCE ADDRESS:
 ADDRESSEE: Quarles & Brady
 STREET: 1 South Pinchney Street
 CITY: Madison
 STATE: WI
 COUNTRY: US

ZIP: 53701-2113
 COMPUTER READABLE FORM:
 MEDIUM TYPE: Floppy disk
 COMPUTER: IBM PC compatible
 OPERATING SYSTEM: PC-DOS/MS-DOS
 SOFTWARE: Patentin Release #1.0, Version #1.30
 CURRENT APPLICATION DATA:
 APPLICATION NUMBER: US/09/056,075
 FILING DATE:

CLASSIFICATION:
 ATTORNEY/AGENT INFORMATION:
 NAME: Seay, Nicholas J.
 REGISTRATION NUMBER: 27386
 REFERENCE/DOCKET NUMBER: 960296.95238
 TELECOMMUNICATION INFORMATION:
 TELEPHONE: 608-251-5000
 TELEFAX: 608-251-9166

INFORMATION FOR SEQ ID NO: 1:
 SEQUENCE CHARACTERISTICS:
 LENGTH: 6243 base pairs
 TYPE: nucleic acid
 STRANDEDNESS: double
 TOPOLOGY: linear
 MOLECULE TYPE: DNA (genomic)
 FEATURE:
 NAME/KEY: misc.feature
 LOCATION: 3770..4013
 OTHER INFORMATION: /note="RP4 origin of DNA transfer (oriT) from
 OTHER INFORMATION: plasmid RP4"

US-09-056-075-1

Query Match 11.4%; Score 40.8; DB 2; Length 6243;

Best Local Similarity 47.3%; Pred. No. 2.3; Indels 1; Gaps 1;

Matches 155; Conservative 0; Mismatches 172; Indels 1; Gaps 1;

4 AAAAAAAGATGAAGATATTTTATTCCTCTTTTGGATTAATTAATCATTT 63
 1416 AATAAAAAAATTTTATTAATTTTATTTTATTTTATTTTATTTTATTTTATTTTATTTTATTTT 1357
 64 TTTTGGCCCAATGATATATAAATTTGGATTAATTAATTAATTAATTAATTTT 122
 1356 TTTTATTAATTTTATTTTATTTTATTTTATTTTATTTTATTTTATTTTATTTTATTTTATTTT 1297
 123 AGTTGGGCTTGAGAAAGGCTTCGACTTCGAAGTGAGCATGATTAATGATGGA 182
 1296 TTTATTTTATTTTATTTTATTTTATTTTATTTTATTTTATTTTATTTTATTTTATTTTATTTT 1237
 183 GCTAGTTGAGCTTTTGGACATTTGATGATGATGATGATGATGATGATGATGAT 242
 1236 TTTTATTAATTTTATTTTATTTTATTTTATTTTATTTTATTTTATTTTATTTTATTTTATTTT 1177
 243 TAACTTAATAGGCTTCTTAATAGGCCAATTAATTAATGATTAATTAACAAAGTGACAA 302
 1176 ACGCTTGAATTAATTAATGATGCTAAGCAAAATTAATGATGCTAAGTAC 1117
 303 CTTTACTGCTTTTGAATCCGACAA 330

Query Match	11.2%	Score 40;	DB 3;	Length 2755;
Best Local Similarity	48.3%;	Pred. No. 3.3;		
Matches 171;	Conservative	0;	Mismatches 180;	Indels 3; Gaps 2
6	AAAAAGATGAAGATTTTATTCCTCTCTTTTTTTTGATGAATTTTAAATCATTTT	65		

Db 2046 ATATACTAGTAAGTATATATTTGACCTTTTAAACGACCTTATGTGGTATTTTCG 2105
Oy 66 TTGGCCCAATGATATATATAAATTTGATTAATATATATGATTTGCTTTTGT 125
Db 2106 TTAATACCAATCAATATAATTTTATTTATTTAATATATGCTACACAGATGTGG 2165
Oy 126 TCGGCTTGGAGAAAAGGTTTGA-CCTTGCAGAGTGCAGATATATAGATTGGAGC 184
Db 2166 TGAAGTAAATATTTTAAACAATATATTTTGAATAATGATTAATAATGAGGT 2225
Oy 185 TAGGTGAGCTTTGGACATTTGTATGTGATGTTTATTTAGTGTGCGACATATA 244
Db 2226 TTGGTGAATAGTATATATATTTTACAAATTTATTTAATATGTAGGTTCAAAATCTATC 2285
Oy 245 AACCTTAATAGGCTTTCTATAGGCCCAATTAATATGATTTATACAAAGTGACACT 304
Db 2286 ATGTGATATTTG--TACTATATATCTATATTAATGATTAACCTTATATAAGTATCTANT 2343
Oy 305 TTACTCTGTTTGTGATCCGAGCAATATACAAATTTGCAATACCAACACACAG 358
Db 2344 TTAGTTATGTTGTTGATGATGATATACCAAAATTTATTAATAATATATTAG 2397

RESULT 11

US-08-341-568-3/C
Sequence 3, Application US/08341568
Patent No. 5661021

GENERAL INFORMATION:

APPLICANT: Buchert, Johanna
APPLICANT: Silka-aho, Matti
APPLICANT: Viikari, Liisa
APPLICANT: Penttila, Merja
APPLICANT: Saloheimo, Anu
APPLICANT: Marjatta, Rannu
TITLE OF INVENTION: Mannanase enzymes, genes coding for them,
TITLE OF INVENTION: methods for isolating the genes, and methods for bleaching 119
NUMBER OF INVENTION: 7
NUMBER OF SEQUENCES: 7
CORRESPONDENCE ADDRESS:
ADDRESSEE: Birch, Stewart, Kolasch and Birch
STREET: PO Box 747
CITY: Falls Church
STATE: VA

COUNTRY: US
ZIP: 22040-0747

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/341,568
FILING DATE: 22-NOV-1994
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: Murphy Jr, Gerald M
REGISTRATION NUMBER: 28,977
REFERENCE/DOCKET NUMBER: 365-262P
TELECOMMUNICATION INFORMATION:
TELEPHONE: (703) 205-8000
TELEFAX: (703) 205-8050
TELEX: 248345

INFORMATION FOR SEQ ID NO: 3:

SEQUENCE CHARACTERISTICS:
LENGTH: 289 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: cDNA
ORIGINAL SOURCE:
ORGANISM: Trichoderma reesei
STRAIN: QM9414

US-08-341-568-3

Query Match

Best Local Similarity 11.0%; Score 39.4; DB 1; Length 289;
Matches 55; Conservative 0; Mismatches 26; Indels 0; Gaps 0;

Oy 20 TATTTTATTCCTTTTGTGATTAATTTAAATCAATTTTGGCCCAATGANA 79
Db 277 TTTTATTTTATTTTATTTTATTTTATTTTATTTTATTTTATTTTATTTGATA 218
Oy 80 TATTAATAATTTGATTAATA 100
Db 217 TATATATATATTCATCAAAA 197

RESULT 12

US-08-911-020-3/C
Sequence 3, Application US/08911020
Patent No. 5854047

GENERAL INFORMATION:

APPLICANT: Buchert, Johanna
APPLICANT: Silka-aho, Matti
APPLICANT: Viikari, Liisa
APPLICANT: Penttila, Merja
APPLICANT: Saloheimo, Anu
APPLICANT: Marjatta, Rannu
TITLE OF INVENTION: Mannanase enzymes, genes coding for them,
TITLE OF INVENTION: methods for isolating the genes, and methods for bleaching
NUMBER OF SEQUENCES: 7
CORRESPONDENCE ADDRESS:
ADDRESSEE: Birch, Stewart, Kolasch and Birch
STREET: PO Box 747
CITY: Falls Church
STATE: VA

COUNTRY: US
ZIP: 22040-0747

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/911,020
FILING DATE: 13-AUG-1997
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/341,568
FILING DATE: 22-NOV-1994
ATTORNEY/AGENT INFORMATION:
NAME: Murphy Jr, Gerald M
REGISTRATION NUMBER: 28,977
REFERENCE/DOCKET NUMBER: 365-262P
TELECOMMUNICATION INFORMATION:
TELEPHONE: (703) 205-8000
TELEFAX: (703) 205-8050
TELEX: 248345

INFORMATION FOR SEQ ID NO: 3:

SEQUENCE CHARACTERISTICS:
LENGTH: 289 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: cDNA
ORIGINAL SOURCE:
ORGANISM: Trichoderma reesei
STRAIN: QM9414
US-08-911-020-3

Query Match

Best Local Similarity 11.0%; Score 39.4; DB 2; Length 289;
Matches 55; Conservative 0; Mismatches 26; Indels 0; Gaps 0;


```

QY      6  AAAAAAGATGAAGTATTTTATTCCTCTTTTGTGATATTTTAAATCATTTT 65
      |||||  |  |||||  |  |||||  |  |||||  |  |||||  |  |||||  |
Db      1513  AAAAAATATTTTGAATTTATTTTATTTTATTAATAATTTTACATCTTTTA 1572
      |||||  |  |||||  |  |||||  |  |||||  |  |||||  |  |||||  |
QY      66  TTGCCCAATGATATATRAAAATTTGGATATAATATATATGATATTCGTTTTAGT 125
      |||||  |  |||||  |  |||||  |  |||||  |  |||||  |  |||||  |
Db      1573  ATACGTTAAATATAAAAAATTTTAAAAAAATTTTAAATATATTTTAAAAAA 1632
      |||||  |  |||||  |  |||||  |  |||||  |  |||||  |  |||||  |
QY      126  TCGGGTTTGAGAAAA 140
      |||||  |  |||||  |  |||||  |  |||||  |  |||||  |  |||||  |
Db      1633  TAATATTTAAAAAAA 1647
      |||||  |  |||||  |  |||||  |  |||||  |  |||||  |  |||||  |

```

Search completed: June 11, 2003, 01:53:51
 Job time : 18.0513 secs

[illegible]

```

RESULT 2
US-10-239-676-140
Sequence 140, Application US/10239676
Publication No. US20030082609A1
GENERAL INFORMATION:
APPLICANT: OLEK, Alexander
APPLICANT: PIEPENBROCK, Christian
APPLICANT: BERLIN, Kurt
TITLE OF INVENTION: Diagnosis of Diseases Associated with Gene Regulation
FILE REFERENCE: 5013.1003
CURRENT APPLICATION NUMBER: US/10/239, 676
CURRENT FILING DATE: 2002-09-24
PRIORITY APPLICATION NUMBER: PCT/EP01/03968
DE 10019058.8
DE 10019173.8
DE 10032529.7
DE 10043626.1
PRIORITY FILING DATE: 2001-04-06
2000-04-06
2000-04-07
2000-06-30
2000-09-01
NUMBER OF SEQ ID NOS: 228
SEQ ID NO 140
LENGTH: 8866
TYPE: DNA
ORGANISM: Artificial Sequence
FEATURE:
OTHER INFORMATION: chemically treated genomic DNA (Homo sapiens)
NAME/KEY: unsure
LOCATION: (3266, 3272, 3300, 3310, 3314..3315, 3334..3335, 3371..3372)
FEATURE:
NAME/KEY: unsure
LOCATION: (3405, 3418)
US-10-239-676-140

```

Query Match	13.9%	Score 49.8	DB 9	Length 8866
Best Local Similarity	49.1%	Pred. No. 1.2		
Matches 132	Conservative 0	Mismatches 137	Indels 0	Gaps 0
QY	22	TTTTATCTCTCTTTTTTTTTTGATTAATTTAAATCAATTTTTTGGCCCAATGATATA	81	
DB	6786	TTTTTATTTGATATTTTTTTTTTAAAGATAGCATTAATTTTTTTATTAATGAATGT	6845	
QY	82	TAAAAATTTGGATTAATATATTTATTTGATATTCGTTTTTACGTGGGCTTGAGAAAG	141	
DB	6846	TTAAATTAAGGAGATTTATTTATGCTAGAGAAATTTTATTTTGTGTGATTTTTGTTGAAT	6905	
QY	142	GGTTTCACCTTTGCCAAAGTGACAGCATGATATATGATTTGGGACGTAGGTTGAGCTTTGCA	201	
DB	6906	TATATTTATTTATATATAGATGTTTTGTTTTAGAAATTTGTTAAATTTAGATTAATTTTATA	6965	
QY	202	CATTTGATTTAGAGATGTTGATATTTAGTGCACACATTTAAACCTTAATAGGGCTTT	261	
DB	6966	GATTTGTTGAATNGAATAATATTTTTTTTAAATATGTATATTTTTTTATATAGGAATA	7025	
QY	262	CTATAGGCCCAATTAATTTACGATTATA	290	
DB	7026	TTAGTTATTAATAATGTTATTTATGTTTAAA	7054	

RESULT 3
US-10-239-676-117

```

Sequence 1//, Application US/102396/6
Publication No. US20030082609A1
GENERAL INFORMATION:
APPLICANT: OLEK, Alexander
APPLICANT: PIENEBROCK, Christian
APPLICANT: BERLIN, Kurt
TITLE OF INVENTION: Diagnosis of Diseases Associated with Gene Regulation
FILE REFERENCE: 5013.1003
CURRENT APPLICATION NUMBER: US/10/239, 676
CURRENT FILING DATE: 2002-09-24
PRIORITY APPLICATION NUMBER: PCT/EP01/03968
DE 10019058.8
DE 10019173.8
DE 10032529.7
DE 10043826.1
PRIORITY FILING DATE: 2001-04-06
2000-04-06
2000-04-07
2000-06-30
2000-09-01
NUMBER OF SEQ ID NOS: 228
SEQ ID NO 117
LENGTH: 11036
TYPE: DNA
ORGANISM: Artificial Sequence
FEATURE:
OTHER INFORMATION: chemically treated genomic DNA (Homo sapiens)
FEATURE:
NAME/KEY: unsure
LOCATION: (1927..1928, 1943, 1956, 1958, 1981)
US-10-239-676-117-

```

Query Match	13.7%	Score 49.2	DB 9	Length 11036
Best Local Similarity	51.4%	Pred. No. 1.6		
Matches 114	Conservative 0	Mismatches 108	Indels 0	Gaps 0
QY	17	AACTATTTTATATCTCTCTCTTTTATTTTGGATAATTTAAACATTTTTTGGCCCAATG	76	
Db	7897	AGCATATTTGATTTTGTATTTTATTTTGGTAATTCGTGACAGTTTATATGTAGCATT	7956	
QY	77	ATATATTTAAATTTGGATTAATATATTTATTTGCATATCTGTTTTTGTTCGGGTTTAC	136	
Db	7957	ATATTTTAATTTATGGAAGAATTTATAGTTTAAAGTTGTTTAAAGACTTTAGGTTATAT	8018	
QY	137	AAAAGGTTTCGACTTTGCAAGAGTGAAGATGTATATAGATTTGGAGCTAGTTAGACT	196	
Db	8017	TTAATTTTATATATTTATATTTTGAAGATTTTGTATTAGTTTATGTATGTTTTAGTTT	8076	
QY	197	TTGAGACTTTGTATGGATGTGTGTATTTATAGTCGCACA	238	
Db	8077	TTGTATATTTTAAATGATGTAGTTTATACATATGTAAATA	8118	

```

RESULT 4
US-10-239-676-36
; Sequence 36, Application US/10239676
; Publication NO. US20030082609A1
; GENERAL INFORMATION:
; APPLICANT: OLEK, Alexander
; APPLICANT: PIEPENBROCK, Christian
; APPLICANT: BERLIN, Kurt
; TITLE OF INVENTION: Diagnosis of Diseases Associated with Gene Regulation
; FILE REFERENCE: 5013 1003
; CURRENT APPLICATION NUMBER: US/10/239, 676
; CURRENT FILING DATE: 2002-09-24
; PRIOR APPLICATION NUMBER: PCT/EP01/03968
; DE 10019058.8
; DE 10019173.8
; DE 10032529.7

```


DE 10043826.1
PRIOR FILING DATE: 2001-04-06
2000-04-06
2000-04-07
2000-06-30
2000-09-01
NUMBER OF SEQ ID NOS: 228
SEQ ID NO 36
LENGTH: 12405
TYPE: DNA
ORGANISM: Artificial Sequence
FEATURE:
OTHER INFORMATION: chemically treated genomic DNA (Homo sapiens)
FEATURE:
NAME/KEY: unsure
LOCATION: (4511)
US-10-239-676-36

Query Match 13.7%; Score 49; DB 9; Length 12405;
Best Local Similarity 47.0%; Pred. No. 1.8;
Matches 151; Conservative 0; Mismatches 170; Indels 0; Gaps 0;

QY 1 AAAAAAAAAAGATGAAATTTTATCTCTCTTTTGTGATATTTAAATCA 60
DB 8896 AAAAAAAAAAGATGAAATTTTATCTCTCTTTTGTGATATTTAAATCA 8955
QY 61 TTTTTCGCCCATGATATATTAATTTGATATATATATATGATATCGTTT 120
DB 8856 TTTTATGTAAGAAATTTGTTTAAATTTTAAATTAATTAATTAATTAAGTAT 9015
QY 121 TTACTGCGGTTTGAGAAAGGTTTGACCTTGAGAGAGATATATAGATTG 180
DB 9016 TGAAGCGATATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 9075
QY 181 GAGCTAGCTGACCTTTGACATTTGATTTGATTTGATTTGATTTGATTTG 240
DB 9076 ATTATAGTATAGATTTTAAATTTTAAATTAATTAATTAATTTTAAATTT 9135
QY 241 ATTAACCTTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 300
DB 9136 AATTAATTTTAAATTAATTAATTTAAATTTGAATTTGAATTTGAATTTGA 9195
QY 301 AACTTACTCTGTTTGTGAT 321
DB 9196 TTGATATTAATTAATTTTAAAT 9216

RESULT 5
US-10-239-676-115
Sequence 115, Application US/10239676
Publication No. US20030082609A1
GENERAL INFORMATION:
APPLICANT: OLEK, Alexander
APPLICANT: PIEPENROCK, Christian
TITLE OF INVENTION: Diagnosis of Diseases Associated with Gene Regulation
FILE REFERENCE: 5013.1003
CURRENT APPLICATION NUMBER: US/10/239,676
PRIOR FILING DATE: 2002-09-24
DE 10019058.8
DE 10019173.8
DE 10032529.7
DE 10043826.1
PRIOR FILING DATE: 2001-04-06
2000-04-06
2000-04-07
2000-06-30
2000-09-01
NUMBER OF SEQ ID NOS: 228
SEQ ID NO 115
LENGTH: 5951
TYPE: DNA

ORGANISM: Artificial Sequence
FEATURE:
OTHER INFORMATION: chemically treated genomic DNA (Homo sapiens)
US-10-239-676-115

Query Match 13.5%; Score 48.4; DB 9; Length 5951;
Best Local Similarity 50.0%; Pred. No. 1.9;
Matches 153; Conservative 0; Mismatches 146; Indels 7; Gaps 1;

QY 20 TATTTTATCTCTCTTTTGTGATATTTAAATCAATTTTGGCCAAATGTA 79
DB 1106 TTTTAAATTTTGTGATTAATTAATTAATTAATTTTGTATTAATTTT 1165
QY 80 TATTAATTTGATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 139
DB 1166 GGTAAATTTGATTTGTTTATTAATTAATTAATTTGTTATTAATTAATTA 1225
QY 140 AGGTTTCGACTTTCGAAAGTGACGATTAATTAATTAATTAATTAATTA 192
DB 1226 TGAATTAATTAATTAATTAATTAATTAATTTTGTGATTTTGTAGTATTA 1285
QY 193 GTCCTTGACATTTGATTAATTAATTTGATTTGATTTGATTTGATTTGAT 252
DB 1286 TGTTTTGACGATTTATGACGTTGATTAATTAATTAATTAATTTTAAATTTG 1345
QY 253 ATGGGCTTCTTAAGGCCCAATTAATTAATTAATTAATTAATTAATTAATTA 312
DB 1346 AGTATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 1405
QY 313 GTTTT 318
DB 1406 TATTTT 1411

RESULT 6
US-10-239-676-198
Sequence 198, Application US/10239676
Publication No. US20030082609A1
GENERAL INFORMATION:
APPLICANT: OLEK, Alexander
APPLICANT: PIEPENROCK, Christian
TITLE OF INVENTION: Diagnosis of Diseases Associated with Gene Regulation
FILE REFERENCE: 5013.1003
CURRENT APPLICATION NUMBER: US/10/239,676
PRIOR FILING DATE: 2002-09-24
DE 10019058.8
DE 10019173.8
DE 10032529.7
DE 10043826.1
PRIOR FILING DATE: 2001-04-06
2000-04-06
2000-04-07
2000-06-30
2000-09-01
NUMBER OF SEQ ID NOS: 228
SEQ ID NO 198
LENGTH: 7823
TYPE: DNA
ORGANISM: Artificial Sequence
FEATURE:
OTHER INFORMATION: chemically treated genomic DNA (Homo sapiens)
US-10-239-676-198

Query Match 13.5%; Score 48.4; DB 9; Length 7823;
Best Local Similarity 46.9%; Pred. No. 2.1;
Matches 151; Conservative 0; Mismatches 171; Indels 0; Gaps 0;

QY 4 AAAAAAAAAAGATGAAATTTTATCTCTCTTTTGTGATATTTAAATCAATTT 63
DB 32 AAAAAAAAAAGATGAAATTTTATCTCTCTTTTGTGATATTTAAATCAATTT 91

Query Match	13.1%;	Score 46.8;	DB 9;	Length 5610;
Best Local Similarity	51.4%;	Pred. No. 3.9;		
Matches 108; Conservative	0;	Mismatches 102;	Indels 0;	Gaps 0;

OY	20	TATTTTAACTCCTCTTTTTTTTTGATTAATTAAATACTTTTTTGCCCAAGTA	79
Db	2137	TAAATTAAATATATATTTTTTTGGAGATGTGTAAATTAATATATATTTTTTATTAATTT	2196
OY	80	TATAAAAATTTGCATTAATAATATTATTCGATATCTTTTTAGTTCCGGTTGACAAA	139
Db	2197	AATTTAATATTTAATGATTTTTTTTTTATTTATTAATTTTAATGAAGTAAATTTGTGAAA	2256
OY	140	AGGCTTCGACTTTCGAAAGTGCACGATGTAATAGATTGGAGCTAGTCTTTG	199
Db	2257	TAAATATTTAAGAATTTATTTGGAAGTAATGAAAAATGTATTTAAGCGAATTAATAG	2316
OY	200	GACATTTGTAATGGATGTGTGATTTTA	229
Db	2317	TATTTTATATATATTTAATGTTTATTTATA	2346

RESULT 10
US-10-239-676-160
; Sequence 160, Application US/10239676
; Publication No. US20030082609A1

```

; APPLICANT: OLEK, Alexander
; APPLICANT: PIEPENBROCK, Christian

```

FILE REFERENCE: 5013 1003

TITLE OF INVENTION: Diagnosis of Diseases Associated with Gene Regulation

DE 10010000 8
PRIOR APPLICATION NUMBER: PCT/EP01/03968
CURRENT FILING DATE: 2002-09-24
CURRENT APPLICATION NUMBER: US/10/239,676

DE 10019058.8
DE 10019173.8
DE 10032529.7
DE 10042000.1

DE 10043826.1
PRIOR FILING DATE: 2001-04-06
2000-04-06
2000-04-06

2000-04-07
2000-06-30
2000-09-01

```

; NUMBER OF SEQ ID NOS: 228
; SEQ ID NO 160
; LENGTH: 9515
;

```

```
;
TYPE: DNA
;
ORGANISM: Artificial Sequence
;
FEATURE:

```

OTHER INFORMATION: chemically treated genomic DNA (Homo sapiens)
US-10-239-676-160

Query match	13.18;	Score 46.8;	DB 9;	Length 9515;
-------------	--------	-------------	-------	--------------

best local similarity 50.4%; Pred. NO. 4.6;
Matches 114; Conservative 0; Mismatches 112; Indels 0; Gaps 0;

5 AAAAAAGATGAAGTATTTTATTTCTCTCTTTTTTTGGATAATTTTAATCATTTTT

1399 AAAAAAAAAAAAAAAAAATATATTCGTTTTTTTAAATAGATGGGAGCTTGT 1455

[illegible]

125 TTCGGGTTTGAGAAAGGGTTTCGACTTTCGAACGTGACGATGTATTAGATTGGACC 184
| | | | | | | | | | | | | | | |
1519 CGTTTTTCGGGGTTTAGATTAGTTTTTTTTTAATTCGAACCTGGGGTTAGATTGGCG 1578

185 TAGGTGAGTCCTTGGACATTTGTATTGGATGTTGTGATTATTAG 230

b 1579 TCGTCTTCTTCTTCTGTTTATATCGTTTATATAATTG 1624

RESULT 11
US-10-239-676-33

GENERAL INFORMATION:
; sequence 33, Application US/102396/8
; Publication No. US20030082609A1

```

; GENERAL INFORMATION.
;
; APPLICANT: OLEK, Alexander
;
; APPLICANT: PIEPENBROCK, Christian
;
; APPLICANT: REBITIN, Kurt

```

FILE REFERENCE: 5013.1003
TITLE OF INVENTION: Diagnosis of Diseases Associated with Gene Regulation
CURRENT APPLICATION NUMBER: US/10/239,676

DE 10019058 8
PRIOR APPLICATION NUMBER: PCT/EP01/03968
CURRENT FILING DATE: 2002-09-24
CORRECTED DISCLOSURE NUMBER: 02/10/2003/01

DE 10043826.1	DE 10032529.7	DE 10019173.8	DE 10015030.0
---------------	---------------	---------------	---------------

PRIOR FILING DATE: 2001-04-06
2000-04-06
2000-04-07

```

; 2000-06-30
; 2000-09-01
NUMBER OF SEO ID NOS: 228
;
;

```

```

; SEQ ID NO 33
; LENGTH: 11735
; TYPE: DNA

```

```

; ORGANISM: Artificial Sequence
;
; FEATURE:
; OTHER INFORMATION: chemically
;

```

OTHER INFORMATION: chemically treated genomic DNA (Homo sapiens)
US-10-239-676-33

Query Match	13.08;	Score 46.6;	DB 9;	Length 11735
Post local similarity	40.88;	Prod No 5.4;		

Best local similarity	43.0%	freq. no. 3.4
Matches 118; Conservative	0;	Mismatches 119; Indels 0; Gaps 0;

18 AGTATTTTATTCCTCTCTTTTGTGATATTTAAACATTTTTTGGCCCAAGA 77

[illegible]

9697 GATAGGTCGTGGGAGGTTTGGCGTTTATTTTCGTTTACGTAGATTTTAAAT 9756

138 AAAGGTTTCGACTTTCGAAGTGACGATGTATATAGATTGGGAGCTAAGTTGAGTCTT 197
||| ||| | ||| | ||| | ||| |
0757 TCACTGTT 0816

198 TGCACATTTGTATTGGATGTTGTTGATTATTAGTGTGACACTATTTAAACCTTAAT 254

9817 AGCATGCTAGTTTGGTATGAGATTGATGAGGTTTCGAGTTTTTAAGTTTAAT 9873

RESULT 12
S-10-239-676-223

Sequence 223, Application US/10239676
Publication No. US20030082609A1
GENERAL INFORMATION:

APPLICANT: OLEK, Alexander
 APPLICANT: PIEPENBROCK, Christian
 APPLICANT: BERLIN, Kurt

TITLE OF INVENTION: Diagnosis of Diseases Associated with Gene Regulation
FILE REFERENCE: 5013.1003
CURRENT APPLICATION NUMBER: US/10/239,676

CURRENT FILING DATE: 2002-09-24
PRIOR APPLICATION NUMBER: PCT/EP01/03968
DE 10019058.8

DE 10019173.8
DE 10032529.7
DE 10043826.1

PRIOR FILING DATE: 2001-04-06
2000-04-06
2000-04-07

[illegible]

DE 10019058.8

DE 10019058.8

DE 10019173.8
DE 10032529.7
DE 10043826.1
PRIOR FILING DATE: 2001-04-06
2000-04-06
2000-04-07
2000-06-30
2000-09-01
NUMBER OF SEQ ID NOS: 228
SEQ ID NO 95
LENGTH: 15732
TYPE: DNA
ORGANISM: Artificial Sequence
FEATURE:
OTHER INFORMATION: chemically treated genomic DNA (Homo sapiens)
US-10-239-676-95

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: June 10, 2003, 22:03:39 ; Search time 860.091 Seconds

(without alignments)
10465.184 Million cell updates/sec

Title: US-09-502-426b-1_COPY_6111_6468

Perfect score: 358
Sequence: 1 aaaaaaaaaagatgaagt.....tgcataataccaacaaga 358

Scoring table:

IDENTITY_NUC
Gap 10.0, Gapext 1.0

Searched: 24791104 seqs, 12571243825 residues

Total number of hits satisfying chosen parameters: 49582208

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 08
Maximum Match 1008

Listing first 45 summaries

Database :

```

1: Pending_Patents_NA_Main:*
2: /cgn2_6/ptodata/1/pna/US06_COMB.seq:*
3: /cgn2_6/ptodata/1/pna/US07_COMB.seq:*
4: /cgn2_6/ptodata/1/pna/US080_COMB.seq:*
5: /cgn2_6/ptodata/1/pna/US081_COMB.seq:*
6: /cgn2_6/ptodata/1/pna/US082_COMB.seq:*
7: /cgn2_6/ptodata/1/pna/US083_COMB.seq:*
8: /cgn2_6/ptodata/1/pna/US084_COMB.seq:*
9: /cgn2_6/ptodata/1/pna/US085_COMB.seq:*
10: /cgn2_6/ptodata/1/pna/US086_COMB.seq:*
11: /cgn2_6/ptodata/1/pna/US087_COMB.seq:*
12: /cgn2_6/ptodata/1/pna/US088_COMB.seq:*
13: /cgn2_6/ptodata/1/pna/US089_COMB.seq:*
14: /cgn2_6/ptodata/1/pna/US090_COMB.seq:*
15: /cgn2_6/ptodata/1/pna/US091_COMB.seq:*
16: /cgn2_6/ptodata/1/pna/US092_COMB.seq:*
17: /cgn2_6/ptodata/1/pna/US093_COMB.seq:*
18: /cgn2_6/ptodata/1/pna/US094_COMB.seq:*
19: /cgn2_6/ptodata/1/pna/US095A_COMB.seq:*
20: /cgn2_6/ptodata/1/pna/US095B_COMB.seq:*
21: /cgn2_6/ptodata/1/pna/US095C_COMB.seq:*
22: /cgn2_6/ptodata/1/pna/US095D_COMB.seq:*
23: /cgn2_6/ptodata/1/pna/US096A_COMB.seq:*
24: /cgn2_6/ptodata/1/pna/US096B_COMB.seq:*
25: /cgn2_6/ptodata/1/pna/US096C_COMB.seq:*
26: /cgn2_6/ptodata/1/pna/US096D_COMB.seq:*
27: /cgn2_6/ptodata/1/pna/US096E_COMB.seq:*
28: /cgn2_6/ptodata/1/pna/US096F_COMB.seq:*
29: /cgn2_6/ptodata/1/pna/US097A_COMB.seq:*
30: /cgn2_6/ptodata/1/pna/US097B_COMB.seq:*
31: /cgn2_6/ptodata/1/pna/US097C_COMB.seq:*
32: /cgn2_6/ptodata/1/pna/US098A_COMB.seq:*
33: /cgn2_6/ptodata/1/pna/US098B_COMB.seq:*
34: /cgn2_6/ptodata/1/pna/US098C_COMB.seq:*
35: /cgn2_6/ptodata/1/pna/US099A_COMB.seq:*
36: /cgn2_6/ptodata/1/pna/US099B_COMB.seq:*
37: /cgn2_6/ptodata/1/pna/US099C_COMB.seq:*
38: /cgn2_6/ptodata/1/pna/US099D_COMB.seq:*
39: /cgn2_6/ptodata/1/pna/US100A_COMB.seq:*
40: /cgn2_6/ptodata/1/pna/US100B_COMB.seq:*
41: /cgn2_6/ptodata/1/pna/US101A_COMB.seq:*
42: /cgn2_6/ptodata/1/pna/US102A_COMB.seq:*
43: /cgn2_6/ptodata/1/pna/US102B_COMB.seq:*

```

Result No.	Score	Query Match	Length	ID	Description
1	358	100.0	6888	1 PCT-US00-03820-1	Sequence 1, Appl1
2	358	100.0	6888	US-09-502-426-1	Sequence 1, Appl1
3	356.4	99.6	745	US-09-505-532-9934	Sequence 9934, Ap
4	356.4	99.6	745	US-09-819-091A-9934	Sequence 9934, Ap
5	356.4	99.6	84196	US-09-692-412-94	Sequence 94, Appl
6	356.4	99.6	84196	US-09-803-736-1074	Sequence 1074, Ap
7	110.4	30.8	566	US-09-654-617-133308	Sequence 133308, Ap
8	110.4	30.8	566	US-09-684-016-133308	Sequence 133308, Ap
9	55.8	15.6	495	US-09-654-617-133308	Sequence 133308, Ap
10	55.8	15.6	495	US-09-684-016-133308	Sequence 133308, Ap
11	54.6	15.3	284	US-09-654-617-17366	Sequence 17366, A
12	54.6	15.3	284	US-09-654-617-339963	Sequence 339963, A
13	54.6	15.3	284	US-09-684-016-339963	Sequence 339963, A
14	54.6	15.3	286	US-09-619-643-14496	Sequence 6235, Ap
15	53.8	15.0	572	US-09-863-418A-12520	Sequence 14496, A
16	53.2	14.9	14147	US-10-172-086-51	Sequence 51, Appl
17	53	14.8	384	US-09-654-617-421464	Sequence 421464, A
18	53	14.8	384	US-09-684-016-421464	Sequence 421464, A
19	53	14.7	384	US-09-197-872-2063	Sequence 2063, Ap
20	52.6	14.7	21231	US-09-614-150-25744	Sequence 25744, A
21	52.6	14.7	21241	US-60-191-637-25874	Sequence 25874, A

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Sequence 20483, A
Sequence 1089, Ap
Sequence 29, APL
Sequence 12690, A
Sequence 18907, A
Sequence 556, App
Sequence 180781, A
Sequence 1607/81, A
Sequence 18013, A
Sequence 339755, A
Sequence 339755, A
Sequence 5695, App
Sequence 367644, A
Sequence 367644, A
Sequence 352823, A
Sequence 165, App
Sequence 61929, A
Sequence 79511, A
Sequence 51282, A
Sequence 1139572, A
Sequence 339872, A
Sequence 6045, App
Sequence 28359, A

UNIVERSITY OF ARIZONA
D USES THEREOF

Sequence	Gap	Position
ATATTAGTGTGACACT	0	240
ATTAGTGTGACACT	0	635
TTATACAAGTAC	0	300
ATACAAGTAC	0	641
TTGGATATTCGTTT	0	120
TTGGATATTCGTTT	0	623
TCGATATTCGTTT	0	180
ATGATATTCGTTT	0	629
ATATATTTAAATCA	0	60
ATATATTTAAATCA	0	617

Sequence	20483, A	QY	301	AAC	TTT	ACT	CG	TTTT	TG	AT	CC	AG	CA	AT	TAC	CA	AA	TT	TC	AA	AT	CC	AA	CA	CA	AG	358
Sequence 1089, Ap1																											
Sequence 29, Ap1																											
Sequence 12650, A																											
		Db	6411	AAC	TTT	ACT	CG	TTTT	TG	AT	CC <td>AG</td> <td>CA</td> <td>AT</td> <td>TAC</td> <td>CA</td> <td>AA</td> <td>TT</td> <td>TC</td> <td>AA</td> <td>AT</td> <td>CC</td> <td>AA</td> <td>CA</td> <td>CA</td> <td>AG</td> <td>6468</td>	AG	CA	AT	TAC	CA	AA	TT	TC	AA	AT	CC	AA	CA	CA	AG	6468

```

Sequence 180781,
Sequence 180781,
Sequence 16013, A
Sequence 339755,
Sequence 339755,
Sequence 5695, Ap
Sequence 36784,
Sequence 36784,
Sequence 35283, A
Sequence 165, App
Sequence 61929, A
Sequence 79531, A
Sequence 51282, A
Sequence 11950, A
Sequence 339872,
Sequence 339872,
Sequence 6045, Ap
Sequence 28359, A

RESULT 2
US-09-502-426-1
; Sequence 1, Application US/09502426
; GENERAL INFORMATION:
; APPLICANT: AZPIROZ, Ricardo
; APPLICANT: CHOE, Sunghwa
; APPLICANT: FELDMANN, Kenneth
; TITLE OF INVENTION: DMP4 POLYNUCLEOTIDES, POLYPEPTIDES AND USES THEREOF
; FILE REFERENCE: 2225-0001
; CURRENT APPLICATION NUMBER: US/09/502,426
; CURRENT FILING DATE: 2000-02-11
; EARLIER APPLICATION NUMBER: 60/119,657
; EARLIER FILING DATE: 1999-02-11
; EARLIER APPLICATION NUMBER: 60/119,658
; EARLIER FILING DATE: 1999-02-11
; NUMBER OF SEQ ID NOS: 18
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 1
; LENGTH: 6888
; TYPE: DNA
; ORGANISM: Arabidopsis sp.
US-09-502-426-1

```

UNIVERSITY OF ARIZONA
 RD USES THEREOF

Query Match 100.0%; Score 358; DB 19; Length 6888;
 Best Local Similarity 100.0%; Pred. No. 9, 2e-44;
 Matches 358; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 AAAAAAAAAAAGATGAAAGATATTTTATATCTCTCTCTTTTCTTTTCTATATTTTAAATCA 60
 DB 6111 AAAAAAAAAAAGATGAAAGATATTTTATATCTCTCTCTTTTCTTTTCTATATTTTAAATCA 6170
 QY 61 TTTTCTTTTGGCCCATGATATATATAAATTTGGATATAATATATTATTTGATATTCGTTT 120
 DB 6171 TTTTCTTTTGGCCCATGATATATATAAATTTGGATATAATATATTATTTGATATTCGTTT 6230
 QY 121 TTAGTTCGGGTTTGAAAGAGGTTTTCGACTTTTGGAAAGTGGACGATATATATAGATTGG 180
 DB 6231 TTAGTTCGGGTTTGAAAGAGGTTTTCGACTTTTGGAAAGTGGACGATATATATAGATTGG 6290
 QY 181 GAGCTAGGTTGAGCTTTTGGACATTTGATTTGATGATGTTGATTATTTAGTGTGACACT 240
 DB 6291 GAGCTAGGTTGAGCTTTTGGACATTTGATTTGATGATGATGTTGATTATTTAGTGTGACACT 6350
 QY 241 ATTAAGCTTAATATGAGCTTCTATAAGAGCCCATTTATTTAGCATATATAACAAATGGAC 300
 DB 6351 ATTAAGCTTAATATGAGCTTCTATAAGAGCCCATTTATTTAGCATATATAACAAATGGAC 6410
 QY 301 AACTTTACTCTGTTTGTGATCCGAGCAATATACAAATTTGCAATATCCAAACACAG 358
 DB 6411 AACTTTACTCTGTTTGTGATCCGAGCAATATACAAATTTGCAATATCCAAACACAG 6468

RESULT 3
 US-09-505-532-9934/C
 ; Sequence 9934, Application US/0950532
 ; GENERAL INFORMATION:
 ; APPLICANT: Cao, Yongwei
 ; APPLICANT: Timberlake, William E.
 ; TITLE OF INVENTION: Plant Genome Sequences and Uses Thereof
 ; FILE REFERENCE: 38-10(15478)B
 ; CURRENT FILING DATE: 2000-02-16
 ; PRIOR APPLICATION NUMBER: USN 60/120,645 1999-02-18; USN 09/443,025 1999-02-18; USN 60/120,645 1999-02-18; USN 09/443,025 1999-11-12
 ; NUMBER OF SEQ ID NOS: 51470
 ; SEQ ID NO 9934
 ; LENGTH: 745

TYPE: DNA
ORGANISM: Arabidopsis thaliana
US-09-502-426b-1
Query Match 99.6%; Score 356.4; DB 19; Length 745;
Best Local Similarity 99.7%; Pred. No. 1.6e-43;
Matches 357; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

1 AAAAAAAAAAGATGATATTTATTTATTTCTCTCTTTTATTTGATTAATTTAAACA 60
632 AAAAAAAAAAGATGATATTTATTTATTTCTCTCTTTTATTTGATTAATTTAAACA 573

61 TTTTTCCTCCCAATGATATATTAATAATTTGATTAATTTGATTAATTTGATTT 120
572 TTTTTCCTCCCAATGATATATTAATAATTTGATTAATTTGATTAATTTGATTT 513

121 TTAGTCGGGTTTGAGAAAAGGCTTTCGACCTTTGCAAGTGAGCATATATAGATTGG 180
512 TTAGTCGGGTTTGAGAAAAGGCTTTCGACCTTTGCAAGTGAGCATATATAGATTGG 453

181 GAGCTAGGTTGAGTCTTTGACATTTGATGATGTTGATTAATTTAGTGCACACT 240
452 GAGCTAGGTTGAGTCTTTGACATTTGATGATGTTGATTAATTTAGTGCACACT 393

241 ATTAACCTTAATGAGGCTTCTATATAGGCCCAATTAATTAATTAATTAATTAAT 300
392 ATTAACCTTAATGAGGCTTCTATATAGGCCCAATTAATTAATTAATTAATTAAT 333

301 AACTTTACTCTGTTTGTGATCCGACCAATTAATTAATTAATTAATTAATTAATTA 358
332 AACTTTACTCTGTTTGTGATCCGACCAATTAATTAATTAATTAATTAATTAATTA 275

RESULT 4

US-09-819-091A-9934/c
Sequence 9934, Application US/09819091A
GENERAL INFORMATION:
APPLICANT: Cao, Yongwei
APPLICANT: Timberlake, William E.
TITLE OF INVENTION: Plant Genome Sequences and Uses Thereof
FILE REFERENCE: 38-10(154)8/B
CURRENT FILING DATE: 2000-02-16
PRIOR APPLICATION NUMBER: US 60/108,420
PRIOR FILING DATE: 1998-11-16
PRIOR APPLICATION NUMBER: US 60/120,645
PRIOR FILING DATE: 1999-02-18
PRIOR APPLICATION NUMBER: US 09/443,025
PRIOR FILING DATE: 1999-11-12
NUMBER OF SEQ ID NOS: 51470
SEQ ID NO 9934
LENGTH: 745
TYPE: DNA
ORGANISM: Arabidopsis thaliana
US-09-819-091A-9934

Query Match 99.6%; Score 356.4; DB 31; Length 745;
Best Local Similarity 99.7%; Pred. No. 1.6e-43;
Matches 357; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

1 AAAAAAAAAAGATGATATTTATTTATTTCTCTCTTTTATTTGATTAATTTAAACA 60
632 AAAAAAAAAAGATGATATTTATTTATTTCTCTCTTTTATTTGATTAATTTAAACA 573

61 TTTTTCCTCCCAATGATATATTAATAATTTGATTAATTTGATTAATTTGATTT 120
572 TTTTTCCTCCCAATGATATATTAATAATTTGATTAATTTGATTAATTTGATTT 513

121 TTAGTCGGGTTTGAGAAAAGGCTTTCGACCTTTGCAAGTGAGCATATATAGATTGG 180
512 TTAGTCGGGTTTGAGAAAAGGCTTTCGACCTTTGCAAGTGAGCATATATAGATTGG 453

181 GAGCTAGGTTGAGTCTTTGACATTTGATGATGTTGATTAATTTAGTGCACACT 240

1 AAAAAAAAAAGATGATATTTATTTATTTCTCTCTTTTATTTGATTAATTTAAACA 60
26092 AAAAAAAAAAGATGATATTTATTTATTTCTCTCTTTTATTTGATTAATTTAAACA 26033

61 TTTTTCCTCCCAATGATATATTAATAATTTGATTAATTTGATTAATTTGATTT 120
26032 TTTTTCCTCCCAATGATATATTAATAATTTGATTAATTTGATTAATTTGATTT 25973

121 TTAGTCGGGTTTGAGAAAAGGCTTTCGACCTTTGCAAGTGAGCATATATAGATTGG 180
25972 TTAGTCGGGTTTGAGAAAAGGCTTTCGACCTTTGCAAGTGAGCATATATAGATTGG 25913

181 GAGCTAGGTTGAGTCTTTGACATTTGATGATGTTGATTAATTTAGTGCACACT 240
25912 GAGCTAGGTTGAGTCTTTGACATTTGATGATGTTGATTAATTTAGTGCACACT 25853

241 ATTAACCTTAATGAGGCTTCTATATAGGCCCAATTAATTAATTAATTAATTAAT 300
25852 ATTAACCTTAATGAGGCTTCTATATAGGCCCAATTAATTAATTAATTAATTAAT 25793

301 AACTTTACTCTGTTTGTGATCCGACCAATTAATTAATTAATTAATTAATTAATTA 358
332 AACTTTACTCTGTTTGTGATCCGACCAATTAATTAATTAATTAATTAATTAATTA 275

RESULT 5

US-09-692-412-94/c
Sequence 94, Application US/09692412
GENERAL INFORMATION:
APPLICANT: Bush, David F.
APPLICANT: Rounsley, Steven D.
APPLICANT: Wiegand, Roger C.
TITLE OF INVENTION: Plant Polymorphic Markers and Uses Thereof
FILE REFERENCE: 38-10(154)93/C
CURRENT FILING DATE: 2000-10-24
NUMBER OF SEQ ID NOS: 124
SEQ ID NO 94
LENGTH: 84196
TYPE: DNA
ORGANISM: Arabidopsis thaliana
US-09-692-412-94

Query Match 99.6%; Score 356.4; DB 27; Length 84196;
Best Local Similarity 99.7%; Pred. No. 1.6e-43;
Matches 357; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

1 AAAAAAAAAAGATGATATTTATTTATTTCTCTCTTTTATTTGATTAATTTAAACA 60
26092 AAAAAAAAAAGATGATATTTATTTATTTCTCTCTTTTATTTGATTAATTTAAACA 26033

61 TTTTTCCTCCCAATGATATATTAATAATTTGATTAATTTGATTAATTTGATTT 120
26032 TTTTTCCTCCCAATGATATATTAATAATTTGATTAATTTGATTAATTTGATTT 25973

121 TTAGTCGGGTTTGAGAAAAGGCTTTCGACCTTTGCAAGTGAGCATATATAGATTGG 180
25972 TTAGTCGGGTTTGAGAAAAGGCTTTCGACCTTTGCAAGTGAGCATATATAGATTGG 25913

181 GAGCTAGGTTGAGTCTTTGACATTTGATGATGTTGATTAATTTAGTGCACACT 240
25912 GAGCTAGGTTGAGTCTTTGACATTTGATGATGTTGATTAATTTAGTGCACACT 25853

241 ATTAACCTTAATGAGGCTTCTATATAGGCCCAATTAATTAATTAATTAATTAAT 300
25852 ATTAACCTTAATGAGGCTTCTATATAGGCCCAATTAATTAATTAATTAATTAAT 25793

301 AACTTTACTCTGTTTGTGATCCGACCAATTAATTAATTAATTAATTAATTAATTA 358
332 AACTTTACTCTGTTTGTGATCCGACCAATTAATTAATTAATTAATTAATTAATTA 275

RESULT 6

US-09-803-736-1074/c
Sequence 1074, Application US/09803736
GENERAL INFORMATION:
APPLICANT: Bush, David F.
APPLICANT: Levin, Irena M.
APPLICANT: Norris, Susan R.
APPLICANT: Rounsley, Steven D.
APPLICANT: Wiegand, Roger C.
TITLE OF INVENTION: Plant Polymorphic Markers and Uses Thereof
FILE REFERENCE: 38-10(154)93/D
CURRENT FILING DATE: 2001-03-12
NUMBER OF SEQ ID NOS: 859
SEQ ID NO 736
LENGTH: 1074
TYPE: DNA
ORGANISM: Arabidopsis thaliana
US-09-803-736-1074/c

PRIOR FILING DATE: 2000-03-29
PRIOR APPLICATION NUMBER: Identified by Attorney Docket number 04983.0206CPUS01 38-10
PRIOR FILING DATE: 2000-10-20
NUMBER OF SEQ ID NOS: 1582
SEQ ID NO 1074
LENGTH: 84196
TYPE: DNA
ORGANISM: Arabidopsis thaliana
US-09-803-736-1074

Query Match
Best Local Similarity 99.6%; Score 356.4; DB 31; Length 84196;
Matches 357; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 AAAAAAAAAAGATGAAAGATATTTATCTCTCTTTTGTGATTAATTAATCA 60
DB 26092 AAAAAAAAAAGATGAAAGATATTTATCTCTCTTTTGTGATTAATTAATCA 26033
QY 61 TTTTGGCCCAATGATATATAAATTTGATTAATTAATTAATTAATCA 120
DB 26032 TTTTGGCCCAATGATATATAAATTTGATTAATTAATTAATTAATCA 25973
QY 121 TTAGTTCGGGTTGAGAAAGGTTTGACCTTCGAAAGTGACGATGATTAATGATGG 180
DB 25972 TTAGTTCGGGTTGAGAAAGGTTTGACCTTCGAAAGTGACGATGATTAATGATGG 25913
QY 181 GACCTAGGTTGAGTCTTTGACATTTTGATGATGTTGATTAATTAATTAATTAATCA 240
DB 25912 GACCTAGGTTGAGTCTTTGACATTTTGATGATGTTGATTAATTAATTAATTAATCA 25853
QY 241 ATTAACCTTAATGGCTTCTTAAGGCCCAATTAATTAATTAATTAATTAATCA 300
DB 25852 ATTAACCTTAATGGCTTCTTAAGGCCCAATTAATTAATTAATTAATTAATCA 25793
QY 301 AACTTCTACTGCTTTTGTGATCCGAAAGCAATTAACAAATTTGCAATTAACAAAGCAG 358
DB 25792 AACTTCTACTGCTTTTGTGATCCGAAAGCAATTAACAAATTTGCAATTAACAAAGCAG 25735

RESULT 7
US-09-617-133308/C
Sequence 133308, Application US/09654617
GENERAL INFORMATION:
APPLICANT: Liu, Jindong
TITLE OF INVENTION: Annotated Plant Genes
FILE REFERENCE: 38-21(15097)D
CURRENT APPLICATION NUMBER: US/09/654,617
CURRENT FILING DATE: 2000-09-05
NUMBER OF SEQ ID NOS: 463173
SEQ ID NO 133308
LENGTH: 566
TYPE: DNA
ORGANISM: Arabidopsis thaliana
OTHER INFORMATION: unsure at all n locations
US-09-654-617-133308

Query Match
Best Local Similarity 90.8%; Score 110.4; DB 25; Length 566;
Matches 128; Conservative 0; Mismatches 12; Indels 1; Gaps 1;

QY 1 AAAAAAAAAAGATGAAAGATATTTATCTCTCTTTTGTGATTAATTAATCA 60
DB 232 AAAAAAAAAAGATGAAAGATATTTATCTCTCTCTTTTGTGATTAATTAATCA 174
QY 61 TTTTGGCCCAATGATATATAAATTTGATTAATTAATTAATTAATCA 120
DB 173 TTTTGGCCCAATGATATATAAATTTGATTAATTAATTAATTAATCA 114
QY 121 TTAGTTCGGGTTGAGAAAG 141
DB 113 TTTNCCCAAAAAAAAAAAG 93

RESULT 8
US-09-684-016-133308/C
Sequence 133308, Application US/09684016
GENERAL INFORMATION:
APPLICANT: Kovalic, David K.
TITLE OF INVENTION: Annotated Plant Genes
FILE REFERENCE: 38-21(15097)D
CURRENT APPLICATION NUMBER: US/09/684,016
CURRENT FILING DATE: 2000-10-10
PRIOR APPLICATION NUMBER: US 09/654,617
PRIOR FILING DATE: 2000-09-05
NUMBER OF SEQ ID NOS: 463173
SEQ ID NO 133308
LENGTH: 566
TYPE: DNA
ORGANISM: Arabidopsis thaliana
FEATURE:
NAME/KEY: unsure
LOCATION: (1)..(566)
OTHER INFORMATION: unsure at all n locations
US-09-684-016-133308

Query Match
Best Local Similarity 90.8%; Score 110.4; DB 27; Length 566;
Matches 128; Conservative 0; Mismatches 12; Indels 1; Gaps 1;

QY 1 AAAAAAAAAAGATGAAAGATATTTATCTCTCTTTTGTGATTAATTAATCA 60
DB 232 AAAAAAAAAAGATGAAAGATATTTATCTCTCTCTTTTGTGATTAATTAATCA 174
QY 61 TTTTGGCCCAATGATATATAAATTTGATTAATTAATTAATTAATCA 120
DB 173 TTTTGGCCCAATGATATATAAATTTGATTAATTAATTAATTAATCA 114
QY 121 TTAGTTCGGGTTGAGAAAG 141
DB 113 TTTNCCCAAAAAAAAAAAG 93

RESULT 9
US-10-021-323-17366/C
Sequence 17366, Application US/10021323
GENERAL INFORMATION:
APPLICANT: Delkman, Jill
TITLE OF INVENTION: Nucleic Acid Molecules and Other Molecules Associated With
FILE REFERENCE: 38-21(52274)B
CURRENT APPLICATION NUMBER: US/10/021,323
CURRENT FILING DATE: 2001-12-12
PRIOR APPLICATION NUMBER: US 60/255, 619
PRIOR FILING DATE: 2000-12-14
NUMBER OF SEQ ID NOS: 17880
SEQ ID NO 17366
LENGTH: 495
TYPE: DNA
ORGANISM: Gossypium hirsutum
FEATURE:
OTHER INFORMATION: Clone ID: LIB3830-001-Q1-N6-B2
US-10-021-323-17366

Query Match
Best Local Similarity 51.4%; Score 55.8; DB 38; Length 495;
Matches 129; Conservative 0; Mismatches 122; Indels 0; Gaps 0;

QY 2 AAAAAAAAAAGATGAAAGATATTTATCTCTCTTTTGTGATTAATTAATCA 61
DB 257 ATAAAGTTAAAGAGGTGACGACATTTAACTTTTTCATGAAAGAAAAAAT 198

```

; RESULT 11
; US-09-654-617-339963/C
; Sequence 339963, Application US/09654617
; GENERAL INFORMATION:
; APPLICANT: Kovalic, David K.
; APPLICANT: Liu, Jingtong
; TITLE OF INVENTION: Annotated Plant Genes
; FILE REFERENCE: 38-21(15097)D
; CURRENT APPLICATION NUMBER: US/09/654, 617
; CURRENT FILING DATE: 2000-09-05

```

```

RESULT 12
US-09-684-016-339963/C
; Sequence 339963, Application US/09684016
; GENERAL INFORMATION:
; APPLICANT: Kowalic, David K.
; APPLICANT: Liu, Jindong
; TITLE OF INVENTION: Annotated Plant Genes
; FILE REFERENCE: 38-21(15097)D
; CURRENT APPLICATION NUMBER: US/09/684, 016
; CURRENT FILING DATE: 2000-10-10
; PRIOR APPLICATION NUMBER: US 09/654,617
; PRIOR FILING DATE: 2000-09-05
; NUMBER OF SEQ ID NOS: 463173
; SEQ ID NO 339963
; LENGTH: 284
; TYPE: DNA
; ORGANISM: Zea mays
US-09-684-016-339963

Query Match          15.3%; Score 54, 6; DB 27; Length 284;
Best Local Similarity 56.4%; Pred. No. 54;
Matches 102; Conservative 0; Mismatches 79; Indels 0; Gaps 0

QY      6 AAAAAAGATGAAGAATTTTATATCTCTCTCTTTTGTTCGATAATTTTAATCATTTT 65
DB      276 ATAAAAAATATATTTTATTTTATTTTATTTTATTTTATTTTATTTTATTTAAT 217
QY      66 TTGGCCATGATATATATAAAATTTGGATAAATATATATGATATTCGTTTAAAT 125
DB      216 TTATATTATTTATTTATTTATTTAAATTTTATTTATTTATTTATTTATTTATTT 157
QY      126 TCGGGTTTGAGAAAAGGTTTCGACTTTCGAAAGTCAGACGATATATAGATTGGAGCT 185
DB      156 TTTTATTTTAAATTAATATAAAATATATAAATATATAAATATAAATAATATATAAAT 97
QY      186 A 186
DB      96 A 96

RESULT 13
US-60-145-485-6295/C
; Sequence 6295, Application US/60145485
; GENERAL INFORMATION:
; APPLICANT: Cheikh, Nordine

```

```

: APPLICANT: Deng, Mollan
: APPLICANT: Fisher, Dane K.
: APPLICANT: Miller, Philip W.
: TITLE OF INVENTION: NOCLETIC ACID MOLECULES AND OTHER MOLECULES ASSOCIATED WITH
: TITLE OF INVENTION: PLANTS
: FILE REFERENCE: 38-21(51229)A
: CURRENT APPLICATION NUMBER: US/60/145,485
: CURRENT FILING DATE: 1999-07-23
: NUMBER OF SEQ ID NOS: 9366
: SEQ ID NO 6295
: LENGTH: 284
: TYPE: DNA
: ORGANISM: Zea mays
: FEATURE:
: OTHER INFORMATION: Clone ID: L1B3150-079-P2-N2-B2
US-60-145-485-6295

```

Query Match	15.3%	Score 54.6	DB 58	Length 284
Best Local Similarity	56.4%	Pred. NO. 54		
Matches 102; Conservative	0	Mismatches 79;	Indels 0;	Gaps 0

QY 6 AAAAAAGTGAAGAATTTTTTTTCCTCTTTTTTTTTTGATAAATTAAACATCTT 65
| | | | | | | | | | | | | | | | | | | | |
Dd 276 ATAAAAAATTAATTTTTTTTATTTTTTTATATTTTTTAATAAAAATTTATATTAATTA 217
| | | | | | | | | | | | | | | | | | | | |
QY 66 TTGCGCAATGATATATATAAAATTTGGATAAATAATATTATGCAATTCGTCTTTTAGT 125
| | | | | | | | | | | | | | | | | | | | |
Dd 216 TTTATATTATTTATTTATTAATTTTTTTTATTTATTTATTTATTTATTTATTTTATTT 157
| | | | | | | | | | | | | | | | | | | | |
QY 126 TCGGTTTGAAGAAAGGCTTTCGACTTCGAAAGTCAGATCATATAGATTGGAGCT 189
| | | | | | | | | | | | | | | | | | | | |
Dd 156 TTTTTTTTTTTAAATTTATATAAATATATAAATTTATTTAAATTTATATAAAT 97
| | | | | | | | | | | | | | | | | | | | |

Qy	186 A 186
	—
Db	96 A 96

```

RESULT 14
US-09-619-643-14496/c
: Sequence 14496, Application US/09619643
: GENERAL INFORMATION:
: APPLICANT: Fisher, Dane K.
: TITLE OF INVENTION: Nucleic Acid Molecules And Other Molecules Associated With
: TITLE OF INVENTION: Plants
: FILE REFERENCE: 38-21(51230)B
: CURRENT APPLICATION NUMBER: US/09/619,643
: CURRENT FILING DATE: 2000-07-19
: NUMBER OF SEQ ID NOS: 32236
: SEQ ID NO 14496
: LENGTH: 286
: TYPE: DNA
: ORGANISM: Zea mays
: OTHER INFORMATION: Clone ID: LTB3150-079-P2-N2-B2
US-09-619-643-14496

```

Query Match	15.3%	Score 54.6	DB 23	Length 286
Best Local Similarity	56.4%	Pred. No. 54		
Matches 102; Conservative	0	Mismatches 79	Indels 0	Gaps 0

[illegible]

QY	186 A 186
Db	98 A 98

```

RESULT 15
US-09-865-419A-12520/c
: Sequence 12520, Application US/09865419A
: GENERAL INFORMATION:
: APPLICANT: Conner, Timothy W.
: APPLICANT: Wu, Kunsheng
: TITLE OF INVENTION: Nucleic Acid Molecules and Other Molecules Associated with
: TITLE OF INVENTION: Plasmids
: FILE REFERENCE: 38-21(51935)B
: CURRENT APPLICATION NUMBER: US/09/865,419A
: PRIORITY FILING DATE: 2001-05-29
: PRIOR APPLICATION NUMBER: US 60/208,063
: PRIOR FILING DATE: 2000-05-31
: NUMBER OF SEQ. ID NOS: 54020
: SEQ. ID NO 12520

```

```

: TYPE: DNA
: ORGANISM: Zea mays
: FEATURE:
: NAME/KEY: unsure
: LOCATION: (1)..(572)
: OTHER INFORMATION: unsure at all n locations
: OTHER INFORMATION: Clone ID: LIB3236-144-PI-N1-H-
: OS-09-865-419A-12510

```

Query Match	15.0%	Score 53.8;	DB 33;	Length 572;
Best Local Similarity	64.8%	Pred. No. 71;		
Matches 79;	Conservative 0;	Mismatches 43;	Indels 0;	Gaps 0

QY	2	AAAAAAAAAGATGAAAGTATTTTATTCCTCTTTTATTTTGGATTAATTTAAATCA	61
Db	392	AAAAAAAAAATTTTAAATTTTATTTTATTTTATTTTATTTTAAAAATTTAAAAAAT	333
QY	62	TTTTTTGGCCCAATGATATATATAAATTTGATATAATATTAATGATATTCGTTTT	121
Db	332	TTTTTTTTTTTTTTTTTTTTTAAAAAATTTTTTTTAAAAAATAATTTTTTTTTTTTT	273
QY			
	122	TA	123
Db	272	TA	271

```
Search completed: June 11, 2003, 05:00:52
Job time : 871.091 secs
```

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OW nucleic - nucleic search, using sw model

Run on: June 10, 2003, 22:22:54 ; Search time 265.216 Seconds
(without alignments)
8960.952 Million cell updates/sec

Title: US-09-502-426B-1_COPY_6111_6468

Perfect score: 358

Sequence: 1 aaaaaaaaaaagatgaagf.....tgcacatcacaacacaaag 358

Scoring table: IDENTITY_NUC
Gapop 10.0, Gapext 1.0

Searched: 7602234 seqs, 3319262570 residues
Total number of hits satisfying chosen parameters: 15204468

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database :

Pending Patents_NA_New:*
1: /cgn2_6/ptodata/1/pna/PCT_NEW_COMB.seq:*
2: /cgn2_6/ptodata/1/pna/US06_NEW_COMB.seq:*
3: /cgn2_6/ptodata/1/pna/US07_NEW_COMB.seq:*
4: /cgn2_6/ptodata/1/pna/US08_NEW_COMB.seq:*
5: /cgn2_6/ptodata/1/pna/US09_NEW_COMB.seq:*
6: /cgn2_6/ptodata/1/pna/US09_NEW_COMB.seq:*
7: /cgn2_6/ptodata/1/pna/US09_NEW_COMB.seq3:*
8: /cgn2_6/ptodata/1/pna/US10_NEW_COMB.seq3:*
9: /cgn2_6/ptodata/1/pna/US10_NEW_COMB.seq2:*
10: /cgn2_6/ptodata/1/pna/US60_NEW_COMB.seq:*
11: /cgn2_6/ptodata/1/pna/US60_NEW_COMB.seq2:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	358	100.0	6888	US-09-502-426B-1	Sequence 1, Appl
2	358	100.0	6888	US-09-502-426A-1	Sequence 1, Appl
3	58	16.2	5449	US-10-311-455-1614	Sequence 1614, Ap
4	58	16.2	7560	US-10-311-455-1196	Sequence 1196, Ap
5	56.4	15.8	5926	US-10-311-455-1625	Sequence 1625, Ap
6	55	15.4	7351	US-10-311-455-1	Sequence 1, Appl
7	55	15.4	20486	US-10-240-453-164	Sequence 164, Appl
8	54.8	15.3	3673778	US-10-312-841-2	Sequence 2, Appl
9	54.2	15.1	2814	US-10-311-455-2100	Sequence 2100, Ap
10	54	15.1	7138	US-10-240-453-329	Sequence 329, Ap
11	53.2	14.9	14147	US-10-311-506-39	Sequence 39, Appl
12	53.2	14.9	14147	US-10-311-507-39	Sequence 39, Appl
13	53.2	14.9	14147	US-10-311-507-39	Sequence 2063, Ap
14	53.2	14.8	384	US-09-837-604A-2063	Sequence 1111, Ap
15	52.8	14.7	5678	US-10-311-455-1111	Sequence 295, App
16	52.2	14.6	6418	US-10-311-455-295	Sequence 494, App
17	52.2	14.6	16750	US-10-311-455-945	Sequence 945, App
18	51.8	14.5	6145	US-10-311-455-1026	Sequence 1026, App
19	51.6	14.4	17131	US-10-311-455-1	Sequence 1, Appl
20	51.2	14.3	3673778	US-10-312-841-1	Sequence 588, Appl
21	51	14.2	13584	US-10-311-455-588	Sequence 992, App
22	50.8	14.2	5311	US-10-311-455-992	Sequence 992, App

23	50.8	14.2	8087	US-10-311-455-715	Sequence 715, App
24	50.8	14.2	13606	US-10-240-453-187	Sequence 187, App
25	50.8	14.2	13606	US-10-311-455-1783	Sequence 1783, App
26	50.6	14.1	5649	US-10-257-166-90	Sequence 90, Appl
27	50.6	14.1	5649	US-10-311-455-822	Sequence 822, App
28	50.4	14.1	5163	US-10-311-455-1221	Sequence 1221, App
29	49.8	13.9	6759	US-10-311-455-711	Sequence 711, App
30	49.8	13.9	8866	US-10-240-453-154	Sequence 154, App
31	49.8	13.9	113515	US-10-311-455-2148	Sequence 2148, App
32	49.6	13.9	9347	US-10-204-708-36	Sequence 36, Appl
33	49.6	13.9	11422	US-10-257-166-17	Sequence 17, Appl
34	49.6	13.9	11422	US-10-311-455-191	Sequence 191, Appl
35	49.6	13.9	13427	US-10-311-455-1899	Sequence 1899, App
36	49.4	13.8	7238	US-10-311-455-421	Sequence 421, App
37	49.4	13.8	8201	US-10-240-452-27	Sequence 27, Appl
38	49.4	13.8	8201	US-10-240-452-27	Sequence 279, App
39	49.4	13.8	9664	US-10-311-455-71	Sequence 71, Appl
40	49.2	13.7	432	US-09-837-604A-67427	Sequence 67427, A
41	49.2	13.7	6464	US-10-311-455-487	Sequence 487, App
42	49.2	13.7	11036	US-10-240-453-137	Sequence 137, App
43	49	13.7	6032	US-10-240-454-6	Sequence 6, Appl
44	49	13.7	6494	US-10-311-455-1366	Sequence 1366, Ap
45	49	13.7	12405	US-10-240-453-44	Sequence 44, Appl

ALIGNMENTS

```

RESULT 1
US-09-502-426B-1
; Sequence 1, Application US/09502426B
; GENERAL INFORMATION:
; APPLICANT: Applioz, Ricardo
; APPLICANT: Choe, Sungjwa
; APPLICANT: Feldmann, Kenneth A.
; TITLE OF INVENTION: DMF POLYNUCLEOTIDES, POLYPEPTIDES AND USES THEREOF
; FILE REFERENCE: 11696-070001
; CURRENT APPLICATION NUMBER: US/09/502,426B
; CURRENT FILING DATE: 2000-02-11
; PRIOR APPLICATION NUMBER: US 60/119,657
; PRIOR FILING DATE: 1999-02-11
; PRIOR APPLICATION NUMBER: US 60/119,658
; PRIOR FILING DATE: 1999-02-11
; NUMBER OF SEQ ID NOS: 30
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 1
; LENGTH: 6888
; TYPE: DNA
; ORGANISM: Arabidopsis thaliana
US-09-502-426B-1

Query Match      100.0%; Score 358; DB 5; Length 6888;
Best Local Similarity 100.0%; Pred. No. 7e-53;
Matches 358; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

1 AAAAAAAAAAAGAGTAAGTATTTATCTCTCTTTTGTGATAATTAATCA 60
|||||
6111 AAAAAAAAAAAGAGTAAGTATTTATCTCTCTTTTGTGATAATTAATCA 6170
|||||
61 TTTTGGCCCAATGATATATAAATTTGATATAATATTAATGATTTG 120
|||||
6171 TTTTGGCCCAATGATATATAAATTTGATATAATATTAATGATTTG 6230
|||||
121 TTAGTGGGTTTGAAGAAAGGTTTGACATTTGCAAGTGCAGATGATATAGATTGG 180.
|||||
TTAGTGGGTTTGAAGAAAGGTTTGACATTTGCAAGTGCAGATGATATAGATTGG 6290
|||||
181 GAGCTAGGTTGAGCTTTGACATTTGATGATGATGATGATTAAGTGCAGACT 240
|||||
6291 GAGCTAGGTTGAGCTTTGACATTTGATGATGATGATGATTAAGTGCAGACT 6350
|||||
241 ATTAACTTAATAGGCTTTCTATTAAGCCCAATTAATTAAGATTAAACAAGTGAC 300
|||||

```

Db 6351 ATTAACCTTAATGGGCTTCTAATAAGCCCAATTATATACATTATTAACAAGTGAC 6410
QY 301 AACTTTACTGCTGTTTGAATCCAGACATTAACAATTTGCAATACCAACAG 358
Db 6411 AACTTTACTGCTGTTTGAATCCAGACATTAACAATTTGCAATACCAACAG 6468

RESULT 2

US-09-502-426a-1
; Sequence 1, Application US/09502426a
; GENERAL INFORMATION:
; APPLICANT: AZPIROZ, Ricardo
; APPLICANT: CHOE, Sunghwa
; APPLICANT: FELDMANN, Kenneth
; TITLE OF INVENTION: DMF4 POLYNUCLEOTIDES, POLYPEPTIDES AND USES THEREOF
; FILE REFERENCE: 2225-0001
; CURRENT APPLICATION NUMBER: US/09/502,426a
; PRIOR FILING DATE: 2000-02-11
; PRIOR APPLICATION NUMBER: 60/119,657
; PRIOR FILING DATE: 1999-02-11
; PRIOR APPLICATION NUMBER: 60/119,658
; PRIOR FILING DATE: 1999-02-11
; NUMBER OF SEQ ID NOS: 25
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 1
; LENGTH: 6888
; TYPE: DNA
; ORGANISM: Arabidopsis sp.
US-09-502-426a-1

Query Match 100.0%; Score 358; DB 7; Length 6888;
Best Local Similarity 100.0%; Pred. No. 7e-53;
Matches 358; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 AAAAAAAAAAAGATGAAGTATTTTATCTCTCTTTTGTGATTAATTTAAATCA 60
Db 6111 AAAAAAAAAAAGATGAAGTATTTTATCTCTCTTTTGTGATTAATTTAAATCA 6170
QY 61 TTTTGGCCCAATGATATATAAATTTGGATTAATATATTTGATATTCGTTT 120
Db 6171 TTTTGGCCCAATGATATATAAATTTGGATTAATATATTTGATATTCGTTT 6230
QY 121 TTGTTGGGTTTGAAGAAAGGTTTGCACCTTGAAGAGTGACATATATGATTGG 180
Db 6231 TTGTTGGGTTTGAAGAAAGGTTTGCACCTTGAAGAGTGACATATATGATTGG 6290
QY 181 GAGCTAGGTGAGCTTTGACATTTGATTTGATTTGATTTAGTGTGACACT 240
Db 6291 GAGCTAGGTGAGCTTTGACATTTGATTTGATTTGATTTAGTGTGACACT 6350
QY 241 ATTAACCTTAATGGGCTTCTAATAAGCCCAATTATATACGATTATAACAAGTGAC 300
Db 6351 ATTAACCTTAATGGGCTTCTAATAAGCCCAATTATATACGATTATAACAAGTGAC 6410
QY 301 AACTTTACTGCTGTTTGAATCCAGACATTAACAATTTGCAATACCAACAG 358
Db 6411 AACTTTACTGCTGTTTGAATCCAGACATTAACAATTTGCAATACCAACAG 6468

RESULT 3

US-10-311-455-1614
; Sequence 1614, Application US/10311455
; GENERAL INFORMATION:
; APPLICANT: OLEK, Alexander
; APPLICANT: PIEPENBROCK, Christian
; APPLICANT: BERLIN, Kurt
; TITLE OF INVENTION: Diagnosis of Diseases Associated with the Immune System by Determining Cytosine Methylation
; FILE REFERENCE: 5013.1014
; CURRENT APPLICATION NUMBER: US/10/311,455
; CURRENT FILING DATE: 2002-12-16
; PRIOR APPLICATION NUMBER: PCT/EP01/07537
; PRIOR FILING DATE: 2001-07-02

; PRIOR APPLICATION NUMBER: DE 10032529.7
; PRIOR FILING DATE: 2000-06-30
; PRIOR APPLICATION NUMBER: DE 10043826.1
; PRIOR FILING DATE: 2000-09-01
; NUMBER OF SEQ ID NOS: 2424
; SEQ ID NO 1614
; LENGTH: 5449
; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: chemically treated genomic DNA (Homo sapiens)
US-10-311-455-1614

Query Match 16.2%; Score 58; DB 9; Length 5449;
Best Local Similarity 65.4%; Pred. No. 0.2;
Matches 85; Conservative 0; Mismatches 45; Indels 0; Gaps 0;

QY 4 AAAAAAAAAAAGATGAAGTATTTTATCTCTCTTTTGTGATTAATTTAAATCA 63
Db 5256 AAAAAAAAAAAGATGAAGTATTTTATCTCTCTTTTGTGATTAATTTAAATCA 5315
QY 64 TTTTGGCCCAATGATATATAAATTTGGATTAATATATTTGATATTCGTTT 123
Db 5316 TGTTTAAGCTTTTATATATATTTTATATATTTTATATTTGATATTTTATTT 5375
QY 124 GTTCGGGTTT 133
Db 5376 TTTTGGGTTT 5385

RESULT 4

US-10-311-455-1196
; Sequence 1196, Application US/10311455
; GENERAL INFORMATION:
; APPLICANT: OLEK, Alexander
; APPLICANT: PIEPENBROCK, Christian
; APPLICANT: BERLIN, Kurt
; TITLE OF INVENTION: Diagnosis of Diseases Associated with the Immune System by Determining Cytosine Methylation
; FILE REFERENCE: 5013.1014
; CURRENT APPLICATION NUMBER: US/10/311,455
; PRIOR FILING DATE: 2002-12-16
; PRIOR APPLICATION NUMBER: PCT/EP01/07537
; PRIOR FILING DATE: 2001-07-02
; PRIOR APPLICATION NUMBER: DE 10032529.7
; PRIOR FILING DATE: 2000-06-30
; PRIOR APPLICATION NUMBER: DE 10043826.1
; PRIOR FILING DATE: 2000-09-01
; NUMBER OF SEQ ID NOS: 2424
; SEQ ID NO 1196
; LENGTH: 7560
; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: chemically treated genomic DNA (Homo sapiens)
US-10-311-455-1196

Query Match 16.2%; Score 58; DB 9; Length 7560;
Best Local Similarity 51.6%; Pred. No. 0.19;
Matches 159; Conservative 0; Mismatches 145; Indels 4; Gaps 1;

QY 2 AAAAAAAAAAAGATGAAGTATTTTATCTCTCTTTTGTGATTAATTTAAATCA 61
Db 2728 AATAATTAACGAATTAAGATTAATTTTATTTTATTTTATTTTATTTATTTA 2787
QY 62 TTTTGGCCCAATGATATATAAATTTGGATTAATATTTGATATTCGTTT 121
Db 2788 TTTTATGCTGAGCTTATTTTAAATTTTAAAGATTAATTTTATTTATGCTGTT 2847
QY 122 TAGTTCGGTGTGAAGAAAGGTTTCGATTCGAAAGTGACGATGATATAGATTGG 181
Db 2848 TGTTTGATATATAAATTTATTTTATGATTTTATTTATTTTATTTTAAATTAAT 2907

QY 182 AGCTAGGTTGAGCTTTGGACATTTGATGATGT----TGTGATTTATAGTGGAC 237
DB 2908 TATATTAATTAATTTGATTAATAGTAATAGTGTGATGATGATTTTAAATTT 2967
QY 228 ACTATTAACCTTAAGGCTTTCTATAGGCCCAATTTATAGCATTAATAGCAACT 297
DB 2968 ATATTTGAATTTAAATAAATTTTATTTTAAAGAAATATATGAAATATATTAAGA 3027
QY 298 GACAACTT 305
DB 3028 GATTAATTT 3035

RESULT 5

US-10-311-455-1625
Sequence 1625, Application US/10311455
GENERAL INFORMATION:
APPLICANT: OLEK, Alexander
APPLICANT: PIEPENBROCK, Christian
APPLICANT: BERLIN, Kurt
TITLE OF INVENTION: Diagnosis of Diseases Associated with the Immune System by Determination of Cytosine Methylation
FILE REFERENCE: 5013.1014
CURRENT APPLICATION NUMBER: US/10/311,455
PRIOR FILING DATE: 2002-12-16
PRIOR APPLICATION NUMBER: PCT/EP01/07537
PRIOR FILING DATE: 2001-07-02
PRIOR APPLICATION NUMBER: DE 10032529.7
PRIOR FILING DATE: 2000-06-30
PRIOR APPLICATION NUMBER: DE 10043826.1
PRIOR FILING DATE: 2000-09-01
NUMBER OF SEQ ID NOS: 2424
SEQ ID NO 1625
LENGTH: 5926
TYPE: DNA
ORGANISM: Artificial Sequence
FEATURE:
OTHER INFORMATION: chemically treated genomic DNA (Homo sapiens)
US-10-311-455-1625

Query Match 15.8%; Score 56.4; DB 9; Length 5926;
Best Local Similarity 48.2%; Pred. No. 0.36;
Matches 159; Conservative 0; Mismatches 171; Indels 0; Gaps 0;

QY 4 AAAAAAAAAAGATGAAGATTTTATCTCTCTTTTATTTTATTTTATTTTAAATTTT 63
DB 3171 AAGCTAGGTTGCGAGTTGATTTGTTTTTTTATTTTATTTTATTTTATTTAGAG 3230
QY 64 TTTTGGCCCAATGATATATTAATAAATTTGATTAATAATATTTATTTGATTTGTTTAA 123
DB 3231 TTTGAAATGTTTATTTATAGAGTTTGTGTTTGGAGTTTGTAGAAATTTAATTTAT 3290
QY 124 GTTCGGGTTTGAGAAAAGGTTTCGACTTTCGAAGTGACATGATATTTAGATGGAG 183
DB 3291 TATGAGATATTTAGTTTATTTTATAGAAAGAGATTTTGAATGAAATATTTTGGGTG 3350
QY 184 CTGAGTTGAGCTTTTGGACATTTGATTTGAGATTTGTTGATTTATTTAGTGCACACTTT 243
DB 3351 GAGGTTTATGATATTTATTTATTTTGTGAGTTATTTGATTAATTTTATTTTATTT 3410
QY 244 AAGCTTAATGAGGTTCTATAGGCCCAATTTATTTATTTATTTATTTAAGTACAC 303
DB 3411 TTAAGGAAAGTTGGTTGGTTGAGAGATTTGGTGGTGGTGGTGAAGTATTTAGTAT 3470
QY 304 TTTTCTGCTTTTGTGATCCGAGCAATTA 333
DB 3471 TATATTTGATTTTGAATGATTTAATAA 3500

RESULT 6

US-10-311-455-1
Sequence 1, Application US/10311455
GENERAL INFORMATION:

APPLICANT: OLEK, Alexander
APPLICANT: PIEPENBROCK, Christian
APPLICANT: BERLIN, Kurt
TITLE OF INVENTION: Diagnosis of Diseases Associated with the Immune System by Determination of Cytosine Methylation
FILE REFERENCE: 5013.1014
CURRENT APPLICATION NUMBER: US/10/311,455
PRIOR FILING DATE: 2002-12-16
PRIOR APPLICATION NUMBER: PCT/EP01/07537
PRIOR FILING DATE: 2001-07-02
PRIOR APPLICATION NUMBER: DE 10032529.7
PRIOR FILING DATE: 2000-06-30
PRIOR APPLICATION NUMBER: DE 10043826.1
PRIOR FILING DATE: 2000-09-01
NUMBER OF SEQ ID NOS: 2424
SEQ ID NO 1
LENGTH: 7351
TYPE: DNA
ORGANISM: Artificial Sequence
FEATURE:
OTHER INFORMATION: chemically treated genomic DNA (Homo sapiens)
US-10-311-455-1

Query Match 15.4%; Score 55; DB 9; Length 7351;
Best Local Similarity 48.0%; Pred. No. 0.61;
Matches 157; Conservative 0; Mismatches 170; Indels 0; Gaps 0;

QY 20 TATTTTATTCCTTTTCTTTTATTTTATTTTAAATTTTAAATTTTATTTTATTTTATTTT 79
DB 3765 TTGTAATTTGTTATGTTATGTTATTTTAAAGATTTTATTTTATTTTATTTTATTTT 3824
QY 80 TATTAATAATTTGATTAATAATTAATTTATTTATTTGATTTGTTTATTTTATTTTATTT 139
DB 3825 ATTAATTTATTTATTTTATTTGATTTATTTATTTTATTTTATTTTATTTTATTTTAA 3884
QY 140 AGGTTTGCACCTTTGAAAGTGACAGATATATAGATTTGAGACCTTGGTGGAGTTG 199
DB 3885 AGCTTTTATGTTTGGTTATGTTTATTTTATTTTATTTTATTTTATTTTATTTTATTT 3944
QY 200 GACATTTATTTGATTTGTTTATTTATTTATTTATTTATTTATTTATTTATTTATTTAT 259
DB 3945 AATAATTTTATTTGTTATTTATTTTATTTATTTATTTATTTATTTATTTATTTATTT 4004
QY 260 TTTTAAAGGCCCAATTTATTTATTTATTTATTTATTTATTTATTTATTTATTTATTT 319
DB 4005 TTTTATTTATTTATTTATTTATTTATTTATTTATTTATTTATTTATTTATTTATTT 4064
QY 320 ATCCGAAGCAATTAACAATTTGCAAT 346
DB 4065 TTTATTAAGAATGTAATAATTTTATAGT 4091

RESULT 7

US-10-240-485-164
Sequence 164, Application US/10240485
GENERAL INFORMATION:
APPLICANT: OLEK, Alexander
APPLICANT: PIEPENBROCK, Christian
APPLICANT: BERLIN, Kurt
TITLE OF INVENTION: Diagnosis of Diseases Associated with the Immune System by Determination of Cytosine Methylation
FILE REFERENCE: 5013.1007
CURRENT APPLICATION NUMBER: US/10/240,485
PRIOR FILING DATE: 2002-10-02
PRIOR APPLICATION NUMBER: PCT/EP01/03970
PRIOR FILING DATE: 2001-04-06
PRIOR APPLICATION NUMBER: DE 10019058.8
PRIOR FILING DATE: 2000-04-06
PRIOR APPLICATION NUMBER: DE 10019173.8
PRIOR FILING DATE: 2000-04-07
PRIOR APPLICATION NUMBER: DE 10032529.7
PRIOR FILING DATE: 2000-06-30
PRIOR APPLICATION NUMBER: DE 10043826.1

PRIOR FILING DATE: 2000-09-01
NUMBER OF SEQ ID NOS: 202
SEQ ID NO 164
LENGTH: 20486
TYPE: DNA
ORGANISM: Artificial Sequence
FEATURE:
OTHER INFORMATION: chemically treated genomic DNA (Homo sapiens)
US-10-240-485-164

Query Match
Best Local Similarity 49.6%; Score 55; DB 9; Length 20486;
Pred. No. 0.52;
Matches 169; Conservative 0; Mismatches 170; Indels 2; Gaps 1;

Db 17 AACGATTTTATTCTCTCTTTTTTTTTTGATAAATTTAAATCATTTTTTGGCCCAATG 76
11836 AGCTACATGCATTTTAATAATTTAGTAATATGTCTTTTAAAGTTTTTTTTAAAAA 11895
QY 77 ATATATAAAAATTTGGATAAATAATATTATGGAATTCGTTTTTAGTCGGTTGAG 136
11896 GTAAACAAGAATTAAGATTAATATATATTATAGATTTTTTACTTTTTTAAGAATTTTTT 11955
QY 137 AAAAGGTTTCGACTTCGAAAGTCGACGATGTAATGATGGAGCTAGGTGAGTCT 196
11956 TTGTTTTCGAATTTTGGTAAGCTTAAGTTTAATATATTATTTTGTGTTTATTTT 12015
Db 197 TTGGCATTGTGATTTGGATGTTGTTGATTTATTTAGTGTGCGACATTTAAACCCTTAATG 256
12016 TTAGCAGTTTATTAAGTATTAGTTTGTAGTAGATGATTTAAATTAATA--TAATTT 12073
QY 257 GCCTTCATTAAGGCCCAATATATTACGATTTATAACAAGTACCACTTTTACTTCGTT 316
12074 GATATAAATGATTTATCGGTAAATAGTATTAAATAAGATTTAAATTTTAAATTTGTTT 12133
QY 317 TTGATCCGAAGCAATTAACAATTTGTCAAATCCAAACACAA 357
Db 12134 TTTATGTTAAAAATATATAATTTATTTAATTTGAAATTTAA 12174

RESULT 8
US-10-312-841-2
Sequence 2. Application US/10312841
GENERAL INFORMATION:
APPLICANT: Epigenomics AG
TITLE OF INVENTION: Diagnose von bedeutenden genetischen Parametern innerhalb des HMC
FILE REFERENCE: E01/1208/MO
CURRENT APPLICATION NUMBER: US/10/312, 841
CURRENT FILING DATE: 2002-12-30
NUMBER OF SEQ ID NOS: 2
SEQ ID NO 2
LENGTH: 3673778
TYPE: DNA
ORGANISM: Artificial Sequence
FEATURE:
OTHER INFORMATION: chemically treated genomic DNA (Homo sapiens)
FEATURE:
NAME/KEY: unsure
LOCATION: (379615)
US-10-312-841-2

Query Match
Best Local Similarity 47.6%; Score 54.8; DB 9; Length 3673778;
Pred. No. 0.26;
Matches 161; Conservative 0; Mismatches 177; Indels 0; Gaps 0;

QY 8 AAAGATGAAGATTTTATTCCTCTCTTTTTTTTTTGATAAATTTAAATCATTTTTTTT 67
Db 2157283 AATGAAATTTTGAAGTATTTTTCGRTGTTTATTTTGTAGCGTTTGTAT 2157342
QY 68 TGCCCAATGATATATAAAATTTGGATAAATAATATTATTTGATTTTCTTTTACTTC 127
Db 2157343 TATTTATATTTAAGTAGAACCTTAGAGTGTGTAGAGATGAGGTGCTGTTTGTGCTTG 2157402
QY 128 GGCTTTGACAAAAGGCTTTCGACTTTCGAAAGTGCAGCATGATATAGATTGGAGCTAG 187

```

Db      2157403 TTTTCTAGTGTAGGGGATTAAGGTGGAGGAGGGTGAACGACGGTTTGTGGGAGAGACAGA 2157462
Qy      188 GTTAGCTTTTGGACATTTGTATTTGGATGTTGTGATTAATAGTGTGACACTATTAAAC 247
        |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
Db      2157463 GTAGAGAAATAATGATTTATTATTGTTATAGTGTCTATTATTTTGGGAAATTTATTAA 2157522
Qy      248 CTTAATAGGCGCTTCTATTAAGGCCCAATATATTTCAGATTATTAACAAGTGCACACTTTT 307
Db      2157523 TATTTTAAGTGTGTTTAAATTTAAAGTATGCTTTTAAAAAATAATATATGTAATTT 2157582
Qy      308 ACTGCTTTTGTGATCCGAAGCAATATTAACAATATGTCAAA 345
Db      2157583 ATTATTTGAGGTAGAAATTTAAAAAATATATATATA 2157620

RESULT 9
US-10-311-455-2100
; Sequence 2100, Application US/10311455
; GENERAL INFORMATION:
; APPLICANT: OLEK, Alexander
; APPLICANT: PIEPENBROCK, Christian
; APPLICANT: BERLIN, Kurt
; APPLICANT:
; TITLE OF INVENTION: Diagnosis of Diseases Associated with the Immune System by De
; TITLE OF INVENTION: cytosine methylation
; FILE REFERENCE: 5013.1014
; CURRENT APPLICATION NUMBER: US/10/311,455
; CURRENT FILING DATE: 2002-12-16
; PRIOR APPLICATION NUMBER: PCT/EP01/07537
; PRIOR FILING DATE: 2001-07-02
; PRIOR APPLICATION NUMBER: DE 10032529.7
; PRIOR FILING DATE: 2000-06-30
; PRIOR APPLICATION NUMBER: DE 10043826.1
; PRIOR FILING DATE: 2000-09-01
; NUMBER OF SEQ ID NOS: 2424
; SEQ ID NO 2100
; LENGTH: 2814
; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: chemically treated genomic DNA (Homo sapiens)
US-10-311-455-2100

Query Match      15.1%; Score 54.2; DB 9; Length 2814;
Best Local Similarity: 48.8%; Pred. No. 0.97;
Matches 146; Conservative 0; Mismatches 153; Indels 0; Gaps 0;

Qy      20 TATTTTATCTCTCTCTTTTCTTTTGGATTAATTTTAAATTCATTTTGTGCCAATGARA 79
        |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
Db      3 TTTGTTTAAATTTAATTTATTTATTTATTAATAATATATATAATATATATATATATATA 62
Qy      80 TATAAAATTTGGATAAATTAATATATATGATATTCGTTTTTACTGCGGTTTGAGANA 139
        |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
Db      63 TACGCTAGGTGGATATATATTTTGTATATATTTTGTATATATATATATATATATATATAT 122
Qy      140 AGGCTTTCGACTTCGAAAGTGAGACGATATATAGATTGGAGCTAGCTGAGCTTTG 199
        |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
Db      123 ATGATATAGTATATATATATATATATATATATATATATATATATATATATATATATAT 182
Qy      200 GACATTTTATATGATGTTGTTGATTAATTAAGTGTGACACTTTTAACTTAAATGGGCT 259
        |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
Db      183 TGTATATATATATATATATATATATATATATATATATATATATATATATATATATAT 242
Qy      260 TTTATATAGGCCCAATTTATTTAGCATTAATAACAAGTACACTTTTACTGCTTTT 318
        |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
Db      243 TTTGGGAAGTTTATATATTTATTTAGATTAATAAGGAGATTTTATTTTATTTT 301

```



```

PRIOR APPLICATION NUMBER: DE 10043826.1
PRIOR FILING DATE: 2000-09-01
NUMBER OF SEQ ID NOS: 132
SEQ ID NO 39
LENGTH: 14147
TYPE: DNA
ORGANISM: Artificial Sequence
FEATURE:
OTHER INFORMATION: chemically treated genomic DNA (Homo sapiens)
US-10-311-506-39

Query Match          14.9%; Score 53.2; DB 9; Length 14147;
Best Local Similarity 48.1%; Pred. No. 1.1; 163; Indels      0; Gaps      0;
Matches 151; Conservative 0; Mismatches 163; Indels 0; Gaps 0;

QY      7      AAAAAGATGAAAGTATTTTATTCCTCTCTTTTATTTGATAATTTTAAACATTTT 66
      |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
Db      10811 AAAAAGTTGGTTAGTTGGTTGCTTTCGTATTTGTTTTTTTTTTCGTTGTTTTTTT 10870

QY      67      TTGCCCAATGATATATPAAAAATTGGATAAATAATATATATATGGAATTCGTTTTAGTT 126
      |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
Db      10871 TTTTTCGAAATTAATGTTAGTTTCTTTTCTTTTATTTTATATGTAATGTAATTTTAGTT 10930

QY      127      CGGGTTTGAGAAAAAGGTTTCGACTTTCGAAACTGACATGTATATTAATTGGAGACTA 186
      |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
Db      10931 TTGGGTTTTTTAATATCGGTTGTCGTTTATTTATTTGCTTTGTGTTTTTAATATTTACGAG 10990

```

Db 10991 TTTAGTCTTTTATTTTCTTTTGGCTTTCTTACGTGTTTTTTTGGCGATATTTT 11050

```

Db      11051 TTTTGTGATTTTATTGAAGAAAGTAAGTCTTTTTTTTGTTGATTTTTTTTTTAATAATTT 11110
              || | | | | | | | | | | | | | | | | | | | | | | | | | | | |
OY      307 TACTTCGTTCGA 320
              - - - - - | | | | |
Db      11111 TTATTAATTAATTGA 11124

RESULT 12
US-10-311-507-39
; Sequence 39, Application US/10311507
GENERAL INFORMATION:
APPLICANT: OLEK, Alexander
APPLICANT: PIEPENBROCK, Christian
APPLICANT: BERLIN, Kurt
TITLE OF INVENTION: Method and nucleic acids for the analysis of astrocytomas
FILE REFERENCE: 5013_1013
CURRENT APPLICATION NUMBER: US/10/311,507
PRIORITY FILING DATE: 2002-12-16
PRIOR APPLICATION NUMBER: PCT/EP01/07538
PRIORITY FILING DATE: 2001-07-02
PRIOR APPLICATION NUMBER: DE 10032529.7
PRIORITY FILING DATE: 2000-06-30
```

```

: NUMBER OF SEQ ID NOS: 136
: SEQ ID NO 39
: LENGTH: 14147
: TYPE: DNA
: ORGANISM: Artificial Sequence
: FEATURE:
: OTHER INFORMATION: chemically treated genomic DNA (Homo sapiens)
US-10-311-507-39

Query Match          14.98; Score 53.2; DB 9; Length 14147;
Best Local Similarity 48.18; Pred. No. 1.1;
Matches 151; Conservative 0; Mismatches 163; Indels 0; Gaps 0;

QY      7  AAAAAGATGAAAGATTTTATTCCTCTCTTTTTCGATTAATTTAAATCATTTT 66
        |||||  |  |||  |  |||  |  |||  |  |||  |  |||  |  |||||
db      10811 AAAAAGTTGGTTAGTTGGTTTCGTAATTTGTTTGTTTTTCGTTGTTT 10870

```


GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: June 10, 2003, 22:04:49 ; Search time 542.722 Seconds

(without alignments)
10683.154 Million cell updates/sec

Title: US-09-502-426b-1_COPY_6111_6468

Perfect score: 358
Sequence: 1 aaaaaaaaaagatgaagt.....tgcaataccaacacaag 358

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 16154066 seqs, 8097743376 residues

Total number of hits satisfying chosen parameters: 32308132

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Minimum Match 08
Maximum Match 100%
Listing first 45 summaries

Database :

```
ESF:*
1: em_estdb:*
2: em_estdb:*
3: em_estdb:*
4: em_estdb:*
5: em_estdb:*
6: em_estdb:*
7: em_estdb:*
8: em_estdb:*
9: em_estdb:*
10: em_estdb:*
11: em_estdb:*
12: em_estdb:*
13: em_estdb:*
14: em_estdb:*
15: em_estdb:*
16: em_estdb:*
17: em_estdb:*
18: em_estdb:*
19: em_estdb:*
20: em_estdb:*
21: em_estdb:*
22: em_estdb:*
23: em_estdb:*
24: em_estdb:*
25: em_estdb:*
26: em_estdb:*
27: em_estdb:*
```

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	137.4	38.4	176	AL762543	Arabidops
2	128.4	35.9	194	AL762544	Arabidops
3	121	33.8	423	AV787618	Arabidops
4	68	19.0	1101	CNS00172	Arabidops
5	65.4	18.3	928	CNS00DKY	Arabidops
6	60	16.8	905	CNS00KHX	Arabidops

C	7	58.2	16.3	1027	17	CNS02r50	AL1212733	Tetradon
C	8	58	16.2	1101	17	CNS016L1	AL106896	Drosophila
C	9	57.8	16.1	961	17	A2691838	AL2691838	EMT156TF
C	10	57	15.9	1101	17	CNS00DEJ4	AL069257	Drosophila
C	11	56.6	15.8	894	17	CNS018BG	AL1091926	Drosophila
C	12	56.6	15.8	1147	17	CNS073CX	AL4273731	clone BA0
C	13	56.4	15.8	1101	17	CNS0029N	AL097997	Drosophila
C	14	56.2	15.7	1101	17	CNS0021J	AL0618106	Tetradon
C	15	56	15.6	949	17	CNS04A1H	AL281906	Tetradon
C	16	55.8	15.5	767	17	CNS00A0X	AL055924	Drosophila
C	17	55.6	15.6	1092	17	CNS020K7	AL175696	Tetradon
C	18	55	15.4	1007	17	CNS006X95	AL19462	T3 end of
C	19	54.8	15.3	1101	17	CNS001FB	AL060732	Drosophila
C	20	54.6	15.3	952	17	CNS006V8	AL065867	Drosophila
C	21	54.6	15.3	1002	17	CNS0185X	AL108927	Drosophila
C	22	54.6	15.3	1121	17	CNS01623	AL106197	Drosophila
C	23	54.4	15.2	423	17	CNS001W9	AL075520	Drosophila
C	24	54.2	15.1	866	17	CNS01161	AL099876	Drosophila
C	25	54.2	15.1	1101	17	CNS00EVL	AL069706	Drosophila
C	26	54	15.1	952	17	CNS016RS	AL107122	Drosophila
C	27	54	15.1	993	9	AL561361	AL561361	AL561361
C	28	54	15.1	1200	17	CNS016CO	AL106578	Drosophila
C	29	53.6	15.0	1043	17	CNS0145P	AL103735	Drosophila
C	30	53.6	15.0	1094	17	CNS012F2	AL101513	Drosophila
C	31	53.6	15.0	1167	17	CNS07360	AL427102	clone BA0
C	32	53.4	14.9	973	17	CNS071LE	AL425064	clone BA0
C	33	53.4	14.9	974	17	AC043196	AC043196	pan troy1
C	34	53.2	14.9	458	9	AL514085	AL514085	AL514085
C	35	53.2	14.9	994	17	CNS0058N	AL057169	Drosophila
C	36	53	14.8	573	17	CNS06LTX	AL046355	T7 end of
C	37	53	14.8	813	17	CNS03CIW	AL237857	Tetradon
C	38	52.8	14.7	637	17	CNS036CC	AL229845	Tetradon
C	39	52.6	14.7	1225	17	CNS0161D	AL106171	Drosophila
C	40	52.4	14.6	529	9	AL514657	AL514657	AL514657
C	41	52.4	14.6	1101	17	CNS0177R	AL107697	Drosophila
C	42	52.4	14.6	1201	17	CNS01038	AL098462	Drosophila
C	43	52.2	14.6	403	17	AC080113	AC080113	HS 4868_B
C	44	52.2	14.6	759	17	CNS060XV	AL411257	T7 end of
C	45	52.2	14.6	765	17	BH148938	BH148938	EMPT52TF

ALIGNMENTS

RESULT 1	AL762543	176 bp	DNA	linear	GSS 19-JUN-2002
LOCUS	AL762543				
DEFINITION	Arabidopsis thaliana T-DNA flanking sequence GK-025G03-013722, genomic survey sequence.				
ACCESSION	AL762543				
VERSION	AL762543.1	GI:21508636			
KEYWORDS	GSS.				
SOURCE	thale cress.				
ORGANISM	Arabidopsis thaliana				
	Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae; eurosids II; Brassicales; Brassicaceae; Arabidopsids.				
REFERENCE	1	Strizhov, N., Li, Y., Rosso, M., Viehoever, P., Dekker, K., Saedler, H. and Weissshaar, B.			
AUTHORS	A pipeline for automated high-throughput generation of FSTS (flanking sequence tags) from Arabidopsis thaliana T-DNA transformed lines				
TITLE	Unpublished				
JOURNAL	2	Rosso, M., Strizhov, N., Li, Y., Reiss, B., Dekker, K. and Weissshaar, B.			
REFERENCE	A new Arabidopsis thaliana T-DNA mutagenised population (GABI-Kat) for flanking sequence tag based reverse genetics				
AUTHORS	3	(bases 1 to 176)			
TITLE	Li, Y., Strizhov, N., Rosso, M. and Weissshaar, B.				
JOURNAL	Direct Submission				
	Submitted (17-JUN-2002) Weissshaar B., Max-Planck-Institut fuer				

COMMENT
Zuechtungsforschung, Carl-von-Linne-Weg 10, Koeln, 50829, Germany
This sequence is recovered from the left border of the T-DNA. It
indicates an insertion within the locus defined by clone T3A5. The
sequences are generated at the MPI for Plant Breeding Research in
the context of the GABI-Kat project. GABI-Kat is part of the German
Plant Genomics program designated 'GABI'. Information on line
availability can be found at:
<http://www.mpiz-koeln.mpg.de/GABI-Kat/>.

FEATURES
source
1. 176
/organism="Arabidopsis thaliana"
/strain="Columbia 0"
/db_xref="taxon:3702"
/clone="GK-025603-013722"
/note="PCR was performed on DNA from Arabidopsis thaliana
plants (T1) which were transformed with the T-DNA from
vector PAC106. The lines contain one or more T-DNA
insertions. The DNA fragment(s) resulting from the PCR
were directly sequenced to determine the genomic sequence
flanking the insertion. Sequences displaying significant
similarity to the A. thaliana nuclear genome sequence were
processed for submission. T-DNA derived sequences were
removed"

BASE COUNT
ORIGIN
60 a 33 c 27 g 56 t

Query Match
Best Local Similarity 38.4%; Score 137.4; DB 17; Length 176;
Matches 138; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 220 TTGATTTTGTGTCGACACTATTAAACCTTAATGGCTTTCATTAAGCCCAATTATA 279
1 TTTGATTTTGTGTCGACACTATTAAACCTTAATGGCTTTCATTAAGCCCAATTATA 60

Db 280 TTACGATTATTAACAAGTGACACTTTTACTCTGTTTTCGCGAAGCAATTAACAATT 339
1 TTTGATTTTGTGTCGACACTATTAAACCTTAATGGCTTTCATTAAGCCCAATTATA 120

QY 61 TTACGATTATTAACAAGTGACACTTTTACTCTGTTTTCGCGAAGCAATTAACAATT 120

QY 340 GTCAATATACCAACACAG 358
121 GTCAATATACCAACACAG 139

Db 121 GTCAATATACCAACACAG 139

RESULT 2
AL762544 194 bp DNA linear GSS 19-JUN-2002
LOCUS Arabidopsis thaliana T-DNA flanking sequence GK-025603-013757,
DEFINITION genomic survey sequence.
ACCESSION AL762544
VERSION AL762544.1 GI:21508641
KEYWORDS GSS.
SOURCE thale cress.
ORGANISM Arabidopsis thaliana
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
Rosidae; eurosids II; Brassicales; Brassicaceae; Arabidopsis.
1
REFERENCE
AUTHORS Strizhov,N., Li,Y., Rosso,M., Vlehoever,P., Dekker,K., Saedler,H.
and Weishaar,B.
TITLE A pipeline for automated high-throughput generation of ESTs
(flanking sequence tags) from Arabidopsis thaliana T-DNA
transformed lines
JOURNAL Unpublished
REFERENCE
AUTHORS Rosso,M., Strizhov,N., Li,Y., Reiss,B., Dekker,K. and Weishaar,B.
TITLE A new Arabidopsis thaliana T-DNA mutagenesis population (GABI-Kat)
for flanking sequence tag based reverse genetics
JOURNAL Unpublished
AUTHORS 3 (bases 1 to 194)
TITLE Submitted (17-JUN-2002) Weishaar B., Max-Planck-Institut fuer

COMMENT
Zuechtungsforschung, Carl-von-Linne-Weg 10, Koeln, 50829, Germany
This sequence is recovered from the left border of the T-DNA. It
indicates an insertion within the locus defined by clone T3A5. The
sequences are generated at the MPI for Plant Breeding Research in
the context of the GABI-Kat project. GABI-Kat is part of the German
Plant Genomics program designated 'GABI'. Information on line
availability can be found at:
<http://www.mpiz-koeln.mpg.de/GABI-Kat/>.

FEATURES
source
1. 194
/organism="Arabidopsis thaliana"
/strain="Columbia 0"
/db_xref="taxon:3702"
/clone="GK-025603-013757"
/note="PCR was performed on DNA from Arabidopsis thaliana
plants (T1) which were transformed with the T-DNA from
vector PAC106. The lines contain one or more T-DNA
insertions. The DNA fragment(s) resulting from the PCR
were directly sequenced to determine the genomic sequence
flanking the insertion. Sequences displaying significant
similarity to the A. thaliana nuclear genome sequence were
processed for submission. T-DNA derived sequences were
removed"

BASE COUNT
ORIGIN
66 a 37 c 28 g 63 t

Query Match
Best Local Similarity 35.9%; Score 128.4; DB 17; Length 194;
Matches 140; Conservative 0; Mismatches 1; Indels 1; Gaps 1;

QY 217 TTGTTGATTATTAGTGTGACACTATTAAACCTTAATGGCTTTCATTAAGCCCAATT 276
27 TTTGTTGATTATTAGTGTGACACT-TTAAACCTTAATGGCTTTCATTAAGCCCAATT 85

Db 277 ATATTAGATTATTAACAAGTGACACTTTTACTCTGTTTTCGCGAAGCAATTAACA 336
86 ATATTAGATTATTAACAAGTGACACTTTTACTCTGTTTTCGCGAAGCAATTAACA 145

QY 337 ATTGTCAATATACCAACACAG 358
146 ATTGTCAATATACCAACACAG 167

Db 146 ATTGTCAATATACCAACACAG 167

RESULT 3
AV787618/c 423 bp mRNA linear EST 28-MAR-2002
LOCUS AV787618 RAE16 Arabidopsis thaliana cDNA clone RAE106-76-F19 3',
DEFINITION mRNA sequence.
ACCESSION AV787618
VERSION AV787618.1 GI:19806408
KEYWORDS EST.
SOURCE thale cress.
ORGANISM Arabidopsis thaliana
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
Rosidae; eurosids II; Brassicales; Brassicaceae; Arabidopsis.
1 (bases 1 to 423)
REFERENCE
AUTHORS Seki,M., Narusawa,M., Ishida,J., Kamiya,A., Satou,M., Nakajima,M.,
Oono,Y., Sakurai,T., Carinli,P., Kawai,J., Itoh,M., Ishii,Y.,
Arakawa,T., Shibata,K., Shingawa,A., Muramatsu,M., Hayashizaki,Y.
and Shinozaki,K.
TITLE Large scale analysis of Arabidopsis full-length cDNA (2002b)
JOURNAL Unpublished (2002)
COMMENT Contact: Motoaki Seki
Plant Functional Genomics Research Group
RIKEN Genomic Sciences Center
3-1-1 Koyadai, Tsukuba, Ibaraki 305-0074, Japan
Tel: 81-298-36-4359
Fax: 81-298-36-9060
Email: mseki@rcc.riken.go.jp
An Arabidopsis full-length cDNA library was constructed essentially
as reported previously (Seki et al., 1998). This clone is in a

```
http://bacpac.med.buffalo.edu/drosophila_bac.htm
Location/Qualifiers
1. 1101
/organism="Drosophila melanogaster"
```

```

/organism="Drosophila melanogaster"
/db_xref="taxon:7227"
/clone="BACR27A24"
/clone_lib="RPCT-98"

```


	Matches	86;	Conservative	0;	Mismatches	47;	Indels	0;	Gaps	0;	
OY		1	AAAAAAAAAAGCAGAATTTTAAATCTCTCTTTTTTTTTTGATAATTAAATCA	60							
Dd		674	AAAAAAAAAAAAATTTTATTTTATTTTATTTTATTTTATTAATAATTAACA	615							
OY		61	TTTTTTTGGCCCATGATATATAAAATTTGGATAAATAATATATGTGATTCGTTT	120							
Dd		614	TTTTTAATTTTATTTTATTAATAAAAATTTAAAAAAATTTTTTTTTTTTTTTTTT	555							
OY		121	TTAGTTCGGGTTT	133							
Dd		554	TTTTTTTTTTTTTT	542							
RESULT 10 CONS00EJ4/C LOCUS		CNS00EJ4	1101 bp	DNA	linear	GSS 04-JUN-1999					
DEFINITION		Drosophila melanogaster genome survey sequence T7 end of BAC:									
ACCESSION		BACR29822 of RPCT-98 library from Drosophila melanogaster (fruit fly), genomic survey sequence.									
VERSION		AL069257									
KEYWORDS		AL069257.1 GI:4949400 GSS.									
SOURCE		Drosophila melanogaster.									
ORGANISM		Drosophila melanogaster. Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha; Epiphytoidae; Drosophilidae; Drosophila.									
REFERENCE		Genoscope. Direct Submission Submitted (02-JUN-1999) Genoscope - Centre National de Sequencage: BP 191 91006 EVRY cedex - FRANCE (E-mail : seqref@genoscope.cns.fr - Web : www.genoscope.cns.fr) determination of this BAC-end sequence was carried out as part of a collaboration with the Berkeley Drosophila Genome Project (BDGP). The BDGP is constructing a physical map of the Drosophila melanogaster genome using these BACs. For further information please see "http://www.fruitfly.org The BDGP Drosophila melanogaster BAC library was prepared by Kazutoyo Osoegawa and Aaron Mamoser in Pieter de Jong's laboratory in the Department of Cancer Genetics at the Roswell Park Cancer Institute in Buffalo, NY. The library is named RPCT-98 and was constructed by partial EcoRI digestion of Drosophila DNA provided by the BDGP from the isogenic strain YZ; cn bw sp, the same strain used for the BDGP's P1 and EST libraries. A more detailed description of the library and how to order individual BAC clones, the entire library, or filters for hybridization from the BACPAC Resource center can be found at http://bacpac.med.buffalo.edu/drosophila_bac.htm.									
COMMENT		Location/Qualifiers									
FEATURES		source	1..1101								
		/organism="Drosophila melanogaster"									
		/db_xref="taxon:7227"									
		/clone="BACR29822"									
		/clone_lib="RPCT-98"									
		/note="end : T7"									
BASE COUNT		434 a	140 c	25 g	377 t	125 others					
ORIGIN											
Query Match		15.9%;	Score 57;	DB 17;	Length 1101;						
Best Local Similarity		42.1%;	Pred. NO. 4.6;								
Matches 149;		Conservative 35;	Mismatches 163;	Indels 7;	Gaps 1;						
OY		1	AAAAAAAAAAGATGAAGATTTTAAATCTCTCTTTTTTTTTTGATAATTAAATCA	60							
Dd		895	AAAMAAATRAATTTTAAAAATTTTATTTTAAATTTTAAAAAATTTWAATTT	836							
OY		61	TTTTTTTGGCCCATGATATATAAAATTTGGATAAATAATATATGTGATTCGTTT	120							
Dd		835	TWATTTTAADADAATAAATAAAGAAAAAAGAAATTTTTCGTAATAAANA	776							
OY		121	TTAGTTCGGGTTT-----GAGAAAGGGTTTCGACTTCGAAAGTGACGATGATAT	173							

DB	Query Match	Best local Similarity	Matches	101: Conservative	44: Mismatches	118: Indels	0: Gaps	0: Indels
Db	775	15.8%	Score 56.6;	DB 17;	Length 894;			
Db	174	38.4%	Pred. No. 6;					
Db	715							
Db	234							
Db	655							
Db	294							
Db	595							
Db	174							
Db	715							
Db	234							
Db	655							
Db	294							
Db	595							
Db	174							
Db	715							
Db	234							
Db	655							
Db	294							
Db	595							
Db	174							
Db	715							
Db	234							
Db	655							
Db	294							
Db	595							
Db	174							
Db	715							
Db	234							
Db	655							
Db	294							
Db	595							
Db	174							
Db	715							
Db	234							
Db	655							
Db	294							
Db	595							
Db	174							
Db	715							
Db	234							
Db	655							
Db	294							
Db	595							
Db	174							
Db	715							
Db	234							
Db	655							
Db	294							
Db	595							
Db	174							
Db	715							
Db	234							
Db	655							
Db	294							
Db	595							
Db	174							
Db	715							
Db	234							
Db	655							
Db	294							
Db	595							
Db	174							
Db	715							
Db	234							
Db	655							
Db	294							
Db	595							
Db	174							
Db</								

[illegible]

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on:

June 10, 2003, 21:51:23 ; Search time 44 Seconds
(without alignments)
1553.581 Million cell updates/sec

Title: US-09-502-426B-2

Perfect score: 2681

Sequence: 1 MFETEHHTLPLLLPSLLS.....FAFPVDFPGLPIRVSRL 513

Scoring table:

BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched:

908470 seqs, 133250620 residues

Total number of hits satisfying chosen parameters:

908470

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing:

Minimum Match 0%
Maximum Match 100%

Listing first 45 summaries

Database :

1: /SIDS2/gcgcdata/geneseq/geneseqp-emb1/AA1980.DAT:*
2: /SIDS2/gcgcdata/geneseq/geneseqp-emb1/AA1981.DAT:*
3: /SIDS2/gcgcdata/geneseq/geneseqp-emb1/AA1982.DAT:*
4: /SIDS2/gcgcdata/geneseq/geneseqp-emb1/AA1983.DAT:*
5: /SIDS2/gcgcdata/geneseq/geneseqp-emb1/AA1984.DAT:*
6: /SIDS2/gcgcdata/geneseq/geneseqp-emb1/AA1985.DAT:*
7: /SIDS2/gcgcdata/geneseq/geneseqp-emb1/AA1986.DAT:*
8: /SIDS2/gcgcdata/geneseq/geneseqp-emb1/AA1987.DAT:*
9: /SIDS2/gcgcdata/geneseq/geneseqp-emb1/AA1988.DAT:*
10: /SIDS2/gcgcdata/geneseq/geneseqp-emb1/AA1989.DAT:*
11: /SIDS2/gcgcdata/geneseq/geneseqp-emb1/AA1990.DAT:*
12: /SIDS2/gcgcdata/geneseq/geneseqp-emb1/AA1991.DAT:*
13: /SIDS2/gcgcdata/geneseq/geneseqp-emb1/AA1992.DAT:*
14: /SIDS2/gcgcdata/geneseq/geneseqp-emb1/AA1993.DAT:*
15: /SIDS2/gcgcdata/geneseq/geneseqp-emb1/AA1994.DAT:*
16: /SIDS2/gcgcdata/geneseq/geneseqp-emb1/AA1995.DAT:*
17: /SIDS2/gcgcdata/geneseq/geneseqp-emb1/AA1996.DAT:*
18: /SIDS2/gcgcdata/geneseq/geneseqp-emb1/AA1997.DAT:*
19: /SIDS2/gcgcdata/geneseq/geneseqp-emb1/AA1998.DAT:*
20: /SIDS2/gcgcdata/geneseq/geneseqp-emb1/AA1999.DAT:*
21: /SIDS2/gcgcdata/geneseq/geneseqp-emb1/AA2000.DAT:*
22: /SIDS2/gcgcdata/geneseq/geneseqp-emb1/AA2001.DAT:*
23: /SIDS2/gcgcdata/geneseq/geneseqp-emb1/AA2002.DAT:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	2681	100.0	513	21	AA07921
2	978.5	36.5	472	18	AAW27153
3	978.5	36.5	472	21	AA644571
4	978.5	36.5	472	21	AA645022
5	978.5	36.5	491	21	AA645021
6	978.5	36.5	492	21	AA644570
7	946	35.3	444	21	AA644572
8	946	35.3	444	21	AA645023
9	866	32.3	512	23	ABB91571
10	816.5	30.3	486	22	AAU02839

11	726	27.1	461	21	AA646490	Arabidopsis thalia
12	726	27.1	462	21	AA646489	Arabidopsis thalia
13	726	27.1	465	21	AA646491	Arabidopsis thalia
14	712	26.6	461	21	AA611836	Arabidopsis thalia
15	712	26.6	462	21	AA611835	Arabidopsis thalia
16	712	26.6	465	21	AA611834	Arabidopsis thalia
17	653.5	24.4	468	21	AA620783	Arabidopsis thalia
18	606	22.6	471	21	AA623013	Arabidopsis thalia
19	606	22.6	479	21	AA623012	Arabidopsis thalia
20	603	22.5	469	21	AA623014	Arabidopsis thalia
21	598.5	22.3	430	21	AA620784	Arabidopsis thalia
22	597.5	22.3	485	22	AAU02835	Arabidopsis thalia
23	596.5	22.2	497	22	AAU02839	Arabidopsis thalia
24	592	22.1	496	22	AAU02837	Arabidopsis thalia
25	588	21.9	509	22	AAU02832	Arabidopsis thalia
26	585	21.8	507	22	AAU02830	Arabidopsis thalia
27	583	21.7	504	22	AAU02822	Arabidopsis thalia
28	571	21.3	501	22	AAU02823	Arabidopsis thalia
29	569.5	21.2	493	22	AAU02834	Arabidopsis thalia
30	568.5	21.2	500	22	AAU02836	Arabidopsis thalia
31	564	21.0	494	22	AAU02824	Arabidopsis thalia
32	564	21.0	500	22	AAU02826	Arabidopsis thalia
33	558.5	20.8	483	22	AAU02827	Arabidopsis thalia
34	558.5	20.8	492	22	AAU02834	Arabidopsis thalia
35	553.5	20.6	488	22	AAU02838	Arabidopsis thalia
36	547.5	20.4	489	21	AA630048	Arabidopsis thalia
37	546	20.4	481	21	AA630049	Arabidopsis thalia
38	534.5	19.9	503	21	AA620785	Arabidopsis thalia
39	528.5	19.7	388	21	AA620785	Arabidopsis thalia
40	503	18.8	492	19	AAW37733	Arabidopsis thalia
41	503	18.8	492	19	AAW41459	Arabidopsis thalia
42	503	18.8	492	22	AAW41459	Arabidopsis thalia
43	503	18.8	492	22	AAW41459	Arabidopsis thalia
44	490	18.3	433	21	AA630050	Arabidopsis thalia
45	461	17.2	497	19	AAW37735	Arabidopsis thalia

ALIGNMENTS

RESULT 1
ID AAB07921 standard; Protein; 513 AA.
XX AAB07921;
AC AAB07921;
DT 14-NOV-2000 (first entry)
DE A cytochrome P450 enzyme designated DMF4.
XX DMF4; cytochrome P450 enzyme; brassinosteroid; 22alpha-hydroxylation;
KW plant phenotype; cell elongation.
XX Arabidopsis sp.
XX WO200047715-A2.
XX PD 17-AUG-2000.
XX PF 11-FEB-2000; 2000WO-US03820.
XX PR 11-FEB-1999; 99US-0119657.
XX PR 11-FEB-1999; 99US-0119658.
XX (ARIZ-) ARIZONA BOARD OF REGENTS.
XX Azpiroz R, Choe S, Feldmann KA;
XX WPI, 2000-549142/50.
XX N-PSDB; AAA59599.
XX New isolated dmf4 polynucleotide useful for altering the phenotype of
XX plants, for diagnostic assays and in the production of antibodies -
XX

XX Claim 50; Fig 11; 113pp; English.

PS The present sequence represents a DMF4 polypeptide. The polypeptide is a

XX cytochrome P450 enzyme that mediates multiple steps in synthesis of

CC brassinosteroids. Specifically, it mediates multiple

CC 22alpha-hydroxylation steps in brassinosteroid biosynthesis. The DMF4

CC polynucleotide is used for altering the phenotype of a plant. DMF4

CC plants display a dramatic reduction in the length of different organs,

CC and this size reduction is attributable to a defect in cell elongation.

CC The DMF4 polynucleotides and polypeptides can be used in diagnostic

CC assays and to generate antibodies, which can be used to produce

CC immunogenic compositions.

XX

XX Sequence 513 AA;

SQ

Query Match 100.0%; Score 2681; DB 21; Length 513;

Best Local Similarity 100.0%; Pred. No. 1.3e-228;

Matches 513; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MFETENHTLLPPLLLPSLLSLFLILKRRNRKTRNLPKSGWPFLEGTIGYLRPYT 60

DB 1 MFETENHTLLPPLLLPSLLSLFLILKRRNRKTRNLPKSGWPFLEGTIGYLRPYT 60

QY 61 ATTLGDMQOQVSKYKTYNSNLFGEPTIYSADAGLNRFTIIONEGRLFECSYPSISGIL 120

DB 61 ATTLGDMQOQVSKYKTYNSNLFGEPTIYSADAGLNRFTIIONEGRLFECSYPSISGIL 120

QY 121 GKSMVLVGDMDHRMDSISLNFISHARLRTILKDYERHTLFYLDSDMOONSIFSADDEA 180

DB 121 GKSMVLVGDMDHRMDSISLNFISHARLRTILKDYERHTLFYLDSDMOONSIFSADDEA 180

QY 181 KKTFFNIMAHNISMDDGEETEQLKKEYVTFMKGVVSAPLNLPGTAYHKAQSRATILK 240

DB 181 KKTFFNIMAHNISMDDGEETEQLKKEYVTFMKGVVSAPLNLPGTAYHKAQSRATILK 240

QY 241 FIERKMERLDTKEEPOEEVEVTEDEAEKMSKSDHNRKQRTDDLLGWLKHSNLTSEQ 300

DB 241 FIERKMERLDTKEEPOEEVEVTEDEAEKMSKSDHNRKQRTDDLLGWLKHSNLTSEQ 300

QY 301 ILDLISLIFAGHETSVAIALAIFFLQACPKAVEELREHELEIARAKKEGESELMWMD 360

DB 301 ILDLISLIFAGHETSVAIALAIFFLQACPKAVEELREHELEIARAKKEGESELMWMD 360

QY 361 YKMDFTQCVINETLRIGNVVRELHRRKALDVRKGYDIPSGWKVLPVISAHLDNSRYD 420

DB 361 YKMDFTQCVINETLRIGNVVRELHRRKALDVRKGYDIPSGWKVLPVISAHLDNSRYD 420

QY 421 QPMLFNFRWQOQNNNGSSGSGSFSTWGNMYMPFGGPRLCAGSELALEMAYFIHHLV 480

DB 421 QPMLFNFRWQOQNNNGSSGSGSFSTWGNMYMPFGGPRLCAGSELALEMAYFIHHLV 480

QY 481 LKFNMLAEDDOPFAFPVDFPNGLPIRVSRIL 513

DB 481 LKFNMLAEDDOPFAFPVDFPNGLPIRVSRIL 513

RESULT 2

AAW27153

ID AAW27153 standard; Protein: 472 AA.

XX AAW27153;

AC

XX 14-APR-1998 (first entry)

DE Arabidopsis thaliana cytochrome P450-type hydroxylase.

XX

XX Arabidopsis thaliana cytochrome P450-type hydroxylase.

KW Cytochrome P450-type hydroxylase; identification; brassinosteroid;

KM brassinosteroid inhibitor; modified plant; recombinant production;

XX

XX Arabidopsis thaliana.

OS

XX

PN WO9735986-A1.

XX

XX 02-OCT-1997.

PD

XX

XX 27-MAR-1997; 97WO-EP01586.

PE

XX

XX 27-MAR-1996; 96US-0622166.

PR

XX

XX (PLAC) MAX PLANCK GES FOERDERUNG WISSENSCHAFTEN.

PA

XX

XX Altmann T, Koncz C, Mathur J, Szekeres MA;

PI

XX

XX WPI: 1997-489649/45.

DR

XX

XX N-PSDB: AAT85306; AAT85307.

DR

XX

XX

PT New isolated plant cytochrome P450-type hydroxylase gene - used to

PT identify substances acting as brassinosteroid(s) or brassinosteroid

PT inhibitors for the production of modified plants

PT

PS Claim 1; Pages 44-46; 77pp; English.

XX

XX The present sequence is Arabidopsis thaliana cytochrome

CC P450-type hydroxylase. The hydroxylase can be used to identify

CC brassinosteroids or brassinosteroid inhibitors, useful to produce

CC plants with modified physiological and/or phenotypic

CC characteristics. The modified plants may show, e.g. stimulation of

CC growth, increased cell elongation, increased wood production,

CC accelerated seed germination at low temperatures, an increase in

CC dry weight, repressed anthocyanin production during growth in light

CC and/or inhibited de-etiolation which is induced, e.g. by cytokinin,

CC in the dark or an increase in stress tolerance. The hydroxylase or

CC its coding sequence can also be used for the recombinant production

CC of compounds, e.g. teasterone.

XX

SQ

Sequence 472 AA;

Query Match 36.5%; Score 978.5; DB 18; Length 472;

Best Local Similarity 41.9%; Pred. No. 6.6e-78;

Matches 210; Conservative 79; Mismatches 171; Indels 41; Gaps 8;

QY 12 LLLPSLSLLFLILKRRNRKTRNLPKSGWPFLEGTIGYLRPYTATTLGDFMOOH 71

DB 7 LLLPSLSLLFLILKRRNRKTRNLPKSGWPFLEGTIGYLRPYTATTLGDFMOOH 71

QY 72 VSKYGIYRSLNLFGEPTIYSADAGLNRFTIIONEGRLFECSYPSISGILGKMSLVLD 131

DB 64 VARGSVYMTLFGEPITFSADPETNRVQLNDEKLFECYSYPASICNLGKHSILMKG 123

QY 132 MHRDMRSISLNFISHARLRTILKDYERHTLFYLDSDMOONSIFSADDEAKKFTYMLAKH 191

DB 124 LHKRMHSITLMSFANSIITKHIMLIDILVRFNLDSSSRVLL--MEBAKKITFELTVKQ 181

QY 192 IMSMDPGEETEQLKKEYVTFMKGVVSAPLNLPGTAYHKAQSRATILKFERKKEERKL 251

DB 182 LMSFDPG-EMSESLKETLLVIEGFSLPLPFTTYKALQAR-----RKAEALVT 232

QY 252 DIKEEPOEEVEVTEDEAEKMSKSDHNRKQRTDDLLGWLKHSNLTSEQIIDLISLFLA 311

DB 233 VVWVKRRREEEGE-----RKDMIALAALDADGSDSEIYDFVALLLVA 278

QY 312 GHETSSVAIALAIFFLQACPKAVEELREHELEIARAKKEGESELMWMDYKMDFTQCVI 371

DB 279 GYETTSITMLAVFLFETPLAQLKEHEKIRAKSD--SYLSEMDYSMPFTQCVV 336

QY 372 NETLRIGNVVRELHRRKALDVRKGYDIPSGWKVLPVISAHLDNSRYDQNLFPMPWQ 431

DB 337 NETLRVANIIGGVFRAMTDVEIRGYKPKWKVSSFRVAHLDPNHRKDAFTFPMWQ 396

QY 432 QONNGASSGSGSFSTWGNMYMPFGGPRLCAGSELALEMAYFIHHLVLFKNMLAED 490

DB 397 -----SNSVTGTGSNVFTFPFGGPRLCAGSELALEMAYFIHHLVGFSGVPAEQ 446

QY 491 DOPFAFPVDFPNGLPIRVSR 511

DB 447 DKLVFFPTTRQKRRPFEVKR 467

RESULT 3
AAG44571
ID AAG44571 standard; Protein: 472 AA.
XX AAG44571;
AC
XX 18-OCT-2000 (first entry)
DT
XX Arabidopsis thaliana protein fragment SEQ ID NO: 55847.
DE
XX Protein identification: signal transduction pathway; metabolic pathway;
KW hybridisation assay; genetic mapping; gene expression control; promoter;
KW termination sequence.
XX
XX Arabidopsis thaliana.
OS
XX EP1033405-A2.
PN
XX 06-SEP-2000.
PD
XX
XX 25-FEB-2000; 2000EP-0301439.
PF
XX
XX 25-FEB-1999; 99US-0121825.
PR 05-MAR-1999; 99US-0123180.
PR 09-MAR-1999; 99US-0123548.
PR 23-MAR-1999; 99US-0125788.
PR 25-MAR-1999; 99US-0126264.
PR 29-MAR-1999; 99US-0126785.
PR 01-APR-1999; 99US-0127462.
PR 06-APR-1999; 99US-0128234.
PR 08-APR-1999; 99US-0128714.
PR 16-APR-1999; 99US-0128845.
PR 19-APR-1999; 99US-0130077.
PR 21-APR-1999; 99US-0130449.
PR 23-APR-1999; 99US-0130510.
PR 23-APR-1999; 99US-0130891.
PR 28-APR-1999; 99US-0131449.
PR 30-APR-1999; 99US-0132048.
PR 04-MAY-1999; 99US-0132407.
PR 05-MAY-1999; 99US-0132484.
PR 06-MAY-1999; 99US-0132485.
PR 06-MAY-1999; 99US-0132486.
PR 07-MAY-1999; 99US-0132487.
PR 11-MAY-1999; 99US-0132863.
PR 14-MAY-1999; 99US-0134256.
PR 14-MAY-1999; 99US-0134218.
PR 14-MAY-1999; 99US-0134219.
PR 14-MAY-1999; 99US-0134221.
PR 14-MAY-1999; 99US-0134370.
PR 18-MAY-1999; 99US-0134768.
PR 19-MAY-1999; 99US-0134941.
PR 20-MAY-1999; 99US-0135124.
PR 21-MAY-1999; 99US-0135353.
PR 24-MAY-1999; 99US-0135629.
PR 25-MAY-1999; 99US-0136021.
PR 27-MAY-1999; 99US-0136392.
PR 28-MAY-1999; 99US-0136782.
PR 01-JUN-1999; 99US-0137228.
PR 03-JUN-1999; 99US-0137528.
PR 04-JUN-1999; 99US-0137502.
PR 07-JUN-1999; 99US-0137724.
PR 08-JUN-1999; 99US-0138094.
PR 10-JUN-1999; 99US-0138540.
PR 10-JUN-1999; 99US-0138847.
PR 14-JUN-1999; 99US-0139119.
PR 16-JUN-1999; 99US-0139452.
PR 16-JUN-1999; 99US-0139453.
PR 17-JUN-1999; 99US-0139492.
PR 18-JUN-1999; 99US-0139454.

PR 18-JUN-1999; 99US-0139455.
PR 18-JUN-1999; 99US-0139456.
PR 18-JUN-1999; 99US-0139457.
PR 18-JUN-1999; 99US-0139458.
PR 18-JUN-1999; 99US-0139459.
PR 18-JUN-1999; 99US-0139460.
PR 18-JUN-1999; 99US-0139461.
PR 18-JUN-1999; 99US-0139462.
PR 18-JUN-1999; 99US-0139463.
PR 18-JUN-1999; 99US-0139750.
PR 18-JUN-1999; 99US-0139763.
PR 21-JUN-1999; 99US-0139817.
PR 22-JUN-1999; 99US-0139899.
PR 23-JUN-1999; 99US-0140353.
PR 23-JUN-1999; 99US-0140354.
PR 24-JUN-1999; 99US-0140695.
PR 28-JUN-1999; 99US-0140823.
PR 29-JUN-1999; 99US-0140991.
PR 30-JUN-1999; 99US-0141287.
PR 01-JUL-1999; 99US-0141842.
PR 01-JUL-1999; 99US-0142154.
PR 02-JUL-1999; 99US-0142055.
PR 06-JUL-1999; 99US-0142390.
PR 08-JUL-1999; 99US-0142803.
PR 09-JUL-1999; 99US-0142920.
PR 12-JUL-1999; 99US-0142977.
PR 13-JUL-1999; 99US-0143542.
PR 14-JUL-1999; 99US-0143624.
PR 15-JUL-1999; 99US-0144005.
PR 16-JUL-1999; 99US-0144085.
PR 16-JUL-1999; 99US-0144086.
PR 19-JUL-1999; 99US-0144325.
PR 19-JUL-1999; 99US-0144331.
PR 19-JUL-1999; 99US-0144332.
PR 19-JUL-1999; 99US-0144333.
PR 19-JUL-1999; 99US-0144334.
PR 19-JUL-1999; 99US-0144335.
PR 20-JUL-1999; 99US-0144352.
PR 20-JUL-1999; 99US-0144632.
PR 20-JUL-1999; 99US-0144884.
PR 21-JUL-1999; 99US-0144814.
PR 21-JUL-1999; 99US-0145086.
PR 21-JUL-1999; 99US-0145087.
PR 22-JUL-1999; 99US-0145088.
PR 22-JUL-1999; 99US-0145089.
PR 22-JUL-1999; 99US-0145192.
PR 23-JUL-1999; 99US-0145145.
PR 23-JUL-1999; 99US-0145218.
PR 23-JUL-1999; 99US-0145224.
PR 26-JUL-1999; 99US-0145276.
PR 27-JUL-1999; 99US-0145913.
PR 27-JUL-1999; 99US-0145918.
PR 27-JUL-1999; 99US-0145919.
PR 28-JUL-1999; 99US-0145951.
PR 02-AUG-1999; 99US-0146386.
PR 02-AUG-1999; 99US-0146388.
PR 02-AUG-1999; 99US-0146389.
PR 03-AUG-1999; 99US-0147038.
PR 04-AUG-1999; 99US-0147204.
PR 04-AUG-1999; 99US-0147302.
PR 05-AUG-1999; 99US-0147192.
PR 05-AUG-1999; 99US-0147260.
PR 06-AUG-1999; 99US-0147303.
PR 06-AUG-1999; 99US-0147416.
PR 09-AUG-1999; 99US-0147493.
PR 10-AUG-1999; 99US-0147935.
PR 11-AUG-1999; 99US-0148171.
PR 11-AUG-1999; 99US-0148319.
PR 12-AUG-1999; 99US-0148341.
PR 13-AUG-1999; 99US-0148565.
PR 13-AUG-1999; 99US-0148684.
PR 16-AUG-1999; 99US-0149368.

PR 17-AUG-1999; 99US-0149175-
PR 18-AUG-1999; 99US-0149426-
PR 20-AUG-1999; 99US-0149722-
PR 20-AUG-1999; 99US-0149723-
PR 20-AUG-1999; 99US-0149929-
PR 23-AUG-1999; 99US-0149902-
PR 23-AUG-1999; 99US-0149930-
PR 25-AUG-1999; 99US-0150566-
PR 26-AUG-1999; 99US-0150884-
PR 27-AUG-1999; 99US-0151065-
PR 27-AUG-1999; 99US-0151066-
PR 30-AUG-1999; 99US-0151080-
PR 31-AUG-1999; 99US-0151303-
PR 01-SEP-1999; 99US-0151930-
PR 07-SEP-1999; 99US-0152363-
PR 10-SEP-1999; 99US-0153070-
PR 13-SEP-1999; 99US-0153758-
PR 15-SEP-1999; 99US-0154018-
PR 16-SEP-1999; 99US-0154039-
PR 20-SEP-1999; 99US-0154779-
PR 22-SEP-1999; 99US-0155139-
PR 23-SEP-1999; 99US-0155486-
PR 24-SEP-1999; 99US-0155659-
PR 28-SEP-1999; 99US-0156458-
PR 29-SEP-1999; 99US-0156596-
PR 04-OCT-1999; 99US-0157117-
PR 05-OCT-1999; 99US-0157753-
PR 06-OCT-1999; 99US-0157865-
PR 07-OCT-1999; 99US-0158029-
PR 08-OCT-1999; 99US-0158232-
PR 12-OCT-1999; 99US-0158359-
PR 13-OCT-1999; 99US-0159293-
PR 13-OCT-1999; 99US-0159294-
PR 13-OCT-1999; 99US-0159295-
PR 14-OCT-1999; 99US-0159329-
PR 14-OCT-1999; 99US-0159330-
PR 14-OCT-1999; 99US-0159331-
PR 14-OCT-1999; 99US-0159637-
PR 14-OCT-1999; 99US-0159638-
PR 18-OCT-1999; 99US-0159584-
PR 21-OCT-1999; 99US-0160741-
PR 21-OCT-1999; 99US-0160767-
PR 21-OCT-1999; 99US-0160768-
PR 21-OCT-1999; 99US-0160770-
PR 21-OCT-1999; 99US-0160814-
PR 21-OCT-1999; 99US-0160815-
PR 22-OCT-1999; 99US-0160980-
PR 22-OCT-1999; 99US-0160981-
PR 22-OCT-1999; 99US-0160989-
PR 25-OCT-1999; 99US-0161404-
PR 25-OCT-1999; 99US-0161405-
PR 25-OCT-1999; 99US-0161406-
PR 26-OCT-1999; 99US-0161359-
PR 26-OCT-1999; 99US-0161360-
PR 26-OCT-1999; 99US-0161361-
PR 26-OCT-1999; 99US-0161920-
PR 28-OCT-1999; 99US-0161992-
PR 28-OCT-1999; 99US-0161993-
PR 29-OCT-1999; 99US-0162142-
Query Match 36.5%; Score 978.5; DB 21; Length 472;
Best Local Similarity 41.9%; Pred. No. 6; Ge-78;
Matches 210; Conservative 79; Mismatches 171; Indels 41; Gaps 8;

QY 12 LLLLPSTLLFLLLKRRNRKTRFVLPPGSGWMPITGIGLKYATATTLGDPMQOH 71
Db 7 LLLLSIAGFL---LLKRTKRRMGCLPPGSLGLPIGTFOUIGAYKTENPEPPTDER 63
QY 72 VSKYKRIYRNLGEPITVSADAGLNKFTLONEGRLFECSYPRSIGILGKWSMLVGD 131
Db 64 VARYGVSFMTLGEPTISADPETNRFLVONEGKLFECYSPIASICNLGKSHLLMKGS 123

QY 132 MHRDMRSTSLNPLSHARLRTILKQVNRHTLFLVDSMOONSIFSNODEAKKFTENLMK 191
Db 124 LKRMHSILTMSFANSIIRKDHLMIDLRLVFNLDLSWMSRVLL--MEBAKITTFLTVKO 181
QY 192 IMSDOPGEETEOAKKEVVTFMKGVSAPLNPGFAYKALOSRATILKFERKMEERL 251
Db 182 LMSFDPG-EMWSLURKELLVIEGFSPLPLFSTTYKALQAR-----RKVAELT 232
QY 252 DIKEEDQEEEVKTEDEAEMSKSDHVRKQRTDDDLGVLKHSNLSSTEQIIDLILSLFA 311
Db 233 VVYMKRREEEGAE-----RKMDLAALLADDGFSDEIDYFVALLVA 278
QY 312 GHETSSVAIALAIFLQACPAVEELREHLEIARAKKELGSELMNDYKKADTQCVI 371
Db 279 GYETSTTMTLAVKFLTETPLALQKEHEKIRAMKSD--SYLSWSDYKSMPTOCVV 336
QY 372 NETRLGVNVRFLHKKALKDVRKGYDIPSGKVLVPSAVHLONSRYDQPLFPMHMO 431
Db 337 NETLRVANTIGVFRRAATDVEIKYKIPGKVFSSRAVHLDPNHEKARTFPMHMO 396
QY 432 QONNGASSGSGSFSTMGNN-YMPGCGPRLCAGSELAKLEMAVFHHLVLFKNMELAD 490
Db 397 -----SNSVTGSGNVFTPRGGPRLCPGYELARVALSVFLHRLVTFGFSVPAEQ 446
QY 491 DQPAFPVDPNGLPIRVSR 511
Db 447 DKLVEFPYTRQKRPPIVKR 467

RESULT 4

AA645022 standard; Protein: 472 AA.

AA645022;

18-OCT-2000 (first entry)

Arabidopsis thaliana protein fragment SEQ ID NO: 56469.

Protein identification; signal transduction pathway; metabolic pathway; hybridisation assay; genetic mapping; gene expression control; promoter;

termination sequence.

Arabidopsis thaliana.

EP1033405-A2.

06-SEP-2000.

25-FEB-2000; 2000EP-0301439.

PR 25-FEB-1999; 99US-0121825-
PR 05-MAR-1999; 99US-0123180-
PR 09-MAR-1999; 99US-0123548-
PR 23-MAR-1999; 99US-0125788-
PR 23-MAR-1999; 99US-0126264-
PR 29-MAR-1999; 99US-0127462-
PR 01-APR-1999; 99US-0128234-
PR 06-APR-1999; 99US-0128234-
PR 08-APR-1999; 99US-0128714-
PR 16-APR-1999; 99US-0129845-
PR 19-APR-1999; 99US-0130077-
PR 21-APR-1999; 99US-0130449-
PR 23-APR-1999; 99US-0130510-
PR 23-APR-1999; 99US-0130891-
PR 28-APR-1999; 99US-0131449-
PR 30-APR-1999; 99US-0132048-
PR 30-APR-1999; 99US-0132407-
PR 04-MAY-1999; 99US-0132484-
PR 05-MAY-1999; 99US-0132485-
PR 06-MAY-1999; 99US-0132486-
PR 06-MAY-1999; 99US-0132487-
PR 07-MAY-1999; 99US-0132863-

PR 11-MAY-1999; 99US-0134256.
PR 14-MAY-1999; 99US-0134218.
PR 14-MAY-1999; 99US-0134219.
PR 14-MAY-1999; 99US-0134221.
PR 14-MAY-1999; 99US-0134370.
PR 18-MAY-1999; 99US-0134768.
PR 19-MAY-1999; 99US-0134941.
PR 20-MAY-1999; 99US-0135124.
PR 21-MAY-1999; 99US-0135353.
PR 24-MAY-1999; 99US-0135629.
PR 25-MAY-1999; 99US-0136021.
PR 27-MAY-1999; 99US-0136382.
PR 28-MAY-1999; 99US-0136782.
PR 01-JUN-1999; 99US-0137222.
PR 03-JUN-1999; 99US-0137528.
PR 04-JUN-1999; 99US-0137502.
PR 07-JUN-1999; 99US-0137724.
PR 08-JUN-1999; 99US-0138094.
PR 10-JUN-1999; 99US-0138540.
PR 14-JUN-1999; 99US-0138847.
PR 16-JUN-1999; 99US-0139119.
PR 16-JUN-1999; 99US-0139452.
PR 17-JUN-1999; 99US-0139453.
PR 18-JUN-1999; 99US-0139454.
PR 18-JUN-1999; 99US-0139455.
PR 18-JUN-1999; 99US-0139456.
PR 18-JUN-1999; 99US-0139457.
PR 18-JUN-1999; 99US-0139458.
PR 18-JUN-1999; 99US-0139459.
PR 18-JUN-1999; 99US-0139460.
PR 18-JUN-1999; 99US-0139461.
PR 18-JUN-1999; 99US-0139462.
PR 18-JUN-1999; 99US-0139463.
PR 18-JUN-1999; 99US-0139750.
PR 21-JUN-1999; 99US-0139763.
PR 22-JUN-1999; 99US-0139817.
PR 23-JUN-1999; 99US-0139819.
PR 23-JUN-1999; 99US-0140353.
PR 24-JUN-1999; 99US-0140354.
PR 28-JUN-1999; 99US-0140695.
PR 29-JUN-1999; 99US-0140823.
PR 30-JUN-1999; 99US-0140991.
PR 01-JUL-1999; 99US-0141287.
PR 01-JUL-1999; 99US-0141842.
PR 02-JUL-1999; 99US-0142154.
PR 06-JUL-1999; 99US-0142055.
PR 08-JUL-1999; 99US-0142390.
PR 09-JUL-1999; 99US-0142803.
PR 12-JUL-1999; 99US-0142920.
PR 13-JUL-1999; 99US-0142977.
PR 14-JUL-1999; 99US-0143542.
PR 15-JUL-1999; 99US-0143624.
PR 16-JUL-1999; 99US-0144005.
PR 16-JUL-1999; 99US-0144085.
PR 16-JUL-1999; 99US-0144086.
PR 19-JUL-1999; 99US-0144325.
PR 19-JUL-1999; 99US-0144327.
PR 19-JUL-1999; 99US-0144331.
PR 19-JUL-1999; 99US-0144332.
PR 19-JUL-1999; 99US-0144333.
PR 19-JUL-1999; 99US-0144334.
PR 19-JUL-1999; 99US-0144335.
PR 20-JUL-1999; 99US-0144352.
PR 20-JUL-1999; 99US-0144632.
PR 20-JUL-1999; 99US-0144637.
PR 21-JUL-1999; 99US-0144884.
PR 21-JUL-1999; 99US-0144814.
PR 21-JUL-1999; 99US-0145086.
PR 22-JUL-1999; 99US-0145088.
PR 22-JUL-1999; 99US-0145085.
PR 22-JUL-1999; 99US-0145087.
PR 22-JUL-1999; 99US-0145089.
PR 22-JUL-1999; 99US-0145192.
PR 23-JUL-1999; 99US-0145145.

PR 23-JUL-1999; 99US-0145218.
PR 23-JUL-1999; 99US-0145224.
PR 26-JUL-1999; 99US-0145276.
PR 27-JUL-1999; 99US-0145813.
PR 27-JUL-1999; 99US-0145818.
PR 28-JUL-1999; 99US-0145919.
PR 28-JUL-1999; 99US-0145951.
PR 02-AUG-1999; 99US-0146386.
PR 02-AUG-1999; 99US-0146388.
PR 03-AUG-1999; 99US-0146389.
PR 04-AUG-1999; 99US-0147038.
PR 04-AUG-1999; 99US-0147204.
PR 05-AUG-1999; 99US-0147302.
PR 05-AUG-1999; 99US-0147192.
PR 06-AUG-1999; 99US-0147260.
PR 06-AUG-1999; 99US-0147303.
PR 09-AUG-1999; 99US-0147416.
PR 09-AUG-1999; 99US-0147493.
PR 10-AUG-1999; 99US-0147935.
PR 11-AUG-1999; 99US-0148171.
PR 12-AUG-1999; 99US-0148341.
PR 13-AUG-1999; 99US-0148565.
PR 16-AUG-1999; 99US-0148684.
PR 17-AUG-1999; 99US-0149368.
PR 18-AUG-1999; 99US-0149175.
PR 20-AUG-1999; 99US-0149426.
PR 20-AUG-1999; 99US-0149722.
PR 20-AUG-1999; 99US-0149723.
PR 23-AUG-1999; 99US-0149929.
PR 23-AUG-1999; 99US-0149902.
PR 23-AUG-1999; 99US-0149930.
PR 23-AUG-1999; 99US-0150566.
PR 26-AUG-1999; 99US-0150884.
PR 27-AUG-1999; 99US-0151065.
PR 27-AUG-1999; 99US-0151066.
PR 27-AUG-1999; 99US-0151080.
PR 30-AUG-1999; 99US-0151302.
PR 31-AUG-1999; 99US-0151303.
PR 01-SEP-1999; 99US-0151438.
PR 07-SEP-1999; 99US-0151930.
PR 10-SEP-1999; 99US-0152363.
PR 13-SEP-1999; 99US-0153070.
PR 13-SEP-1999; 99US-0153758.
PR 15-SEP-1999; 99US-0154018.
PR 16-SEP-1999; 99US-0154039.
PR 20-SEP-1999; 99US-0154779.
PR 22-SEP-1999; 99US-0155139.
PR 23-SEP-1999; 99US-0155486.
PR 24-SEP-1999; 99US-0155659.
PR 28-SEP-1999; 99US-0156458.
PR 29-SEP-1999; 99US-0156596.
PR 04-OCT-1999; 99US-0157117.
PR 05-OCT-1999; 99US-0157717.
PR 06-OCT-1999; 99US-0157765.
PR 07-OCT-1999; 99US-0158029.
PR 08-OCT-1999; 99US-0158232.
PR 12-OCT-1999; 99US-0158369.
PR 13-OCT-1999; 99US-0159283.
PR 13-OCT-1999; 99US-0159284.
PR 13-OCT-1999; 99US-0159294.
PR 14-OCT-1999; 99US-0159329.
PR 14-OCT-1999; 99US-0159330.
PR 14-OCT-1999; 99US-0159331.
PR 14-OCT-1999; 99US-0159637.
PR 14-OCT-1999; 99US-0159638.
PR 18-OCT-1999; 99US-0159584.
PR 21-OCT-1999; 99US-0160741.
PR 21-OCT-1999; 99US-0160767.
PR 21-OCT-1999; 99US-0160768.
PR 21-OCT-1999; 99US-0160770.
PR 21-OCT-1999; 99US-0160814.
PR 21-OCT-1999; 99US-0160815.
PR 22-OCT-1999; 99US-0160980.

PR 22-OCT-1999; 99US-0160981.
PR 22-OCT-1999; 99US-0160989.
PR 25-OCT-1999; 99US-0161404.
PR 25-OCT-1999; 99US-0161405.
PR 25-OCT-1999; 99US-0161406.
PR 25-OCT-1999; 99US-0161359.
PR 26-OCT-1999; 99US-0161360.
PR 26-OCT-1999; 99US-0161361.
PR 28-OCT-1999; 99US-0161920.
PR 28-OCT-1999; 99US-0161992.
PR 28-OCT-1999; 99US-0161993.
PR 28-OCT-1999; 99US-0162142.

Query Match 36.5%; Score 978.5; DB 21; Length 472;
Best Local Similarity 41.9%; Pred. No. 6.6e-78;
Matches 210; Conservative 79; Mismatches 171; Indels 41; Gaps 8;

QY 12 LLLPLSLLLFLILIKRNRKTRFNLPGKSGMPPLGFTIGLYAKTYTTLTADENQOH 71
DB 7 LLLLSIAAGFL--LLRTRRYRMGLPGSLGLPLIGFTFOLIGAYKTENPEPFLDER 63
QY 72 VSKYGIYRNLGPEPTIVSADAGLNRFILQNGRLEEGSYPRSIGILGKMSMLVVG 131
DB 64 VARYGSYEMTHLGEPTIFESADPETNRFLVQNEGKLEFECYSPASICMLLKRHSLLMKGS 123
QY 132 MHRDMSISINLSHARLTILKLDVERHTLFYLDMSQONSIFSADEAKKFTFNMAKH 191
DB 124 LHRHMSLTFMSFANSSIIKDHMLDRLVFNLDSSSHVLL--MEAKKTFPELVKQ 181
QY 192 IMSMDGEETELQKRYTFMGVSAFLNPGTAHKLQSNATILKTERMERKRL 251
DB 182 LMSFDPG-EMSESLRKEEYLVIGFSLPLRSTYTRKAIQA-----RKVAEALT 232
QY 252 DIKEDEEEETEDAEKMSKDHVKQRTDILLGVKHSNSTEOTILLDLISLFA 311
DB 233 VVYAKRREEEAE-----RKQMLALLAADGFSDEIVDLVALLVA 278
QY 312 GHTSSVAIALAIFLQACKAVEELREHLEIARAKKEIGESELNMDYKKMDFTQCVT 371
DB 279 GYETSTIMLAVKFLTEPLALAOLEKEHEKIRAMKSD--SYLSLESDYKSMPTQCV 336
QY 372 NETLRGNVYRFLHRAKALQVRKYGDIPSGKVLPIYSVHNDNSYDQNLFPNRMQ 431
DB 337 NETLRVANIIGVFRAMTDVEIKTKIPGKVVFSFRAVHDPNFKDARTFNPWRMQ 366
QY 432 QONNGASSSGSFTWGN--YMPFGSGPRLCAGSELAKEMAVFIHVLKFNWELAE 490
DB 397 -----SNSVTIGPSNVFTFPFGGPRLCGYELARVALSVFLHRLVGTGSWPAEQ 446
QY 491 DQFFARPEVDFPGLRVS 511
DB 447 DKLVFPPTRTOKRYPFVKR 467

RESULT 5

AAG45021

ID AAG45021 standard; Protein; 491 AA.

AC AAG45021;

DT 18-OCT-2000 (first entry)

DE Arabidopsis thaliana protein fragment SEQ ID NO: 56468.

XX Protein identification: signal transduction pathway; metabolic pathway;
KM hybridisation assay; genetic mapping; gene expression control; promoter;
KM termination sequence.

XX Arabidopsis thaliana.

XX EP1033405-A2.

PD 06-SEP-2000.

XX 25-FEB-2000; 2000EP-0301439.
PR 25-FEB-1999; 99US-0121825.
PR 05-MAR-1999; 99US-0123180.
PR 09-MAR-1999; 99US-0123348.
PR 23-MAR-1999; 99US-0125788.
PR 25-MAR-1999; 99US-0126264.
PR 29-MAR-1999; 99US-0126785.
PR 01-APR-1999; 99US-0127462.
PR 06-APR-1999; 99US-0128234.
PR 08-APR-1999; 99US-0128714.
PR 16-APR-1999; 99US-0129845.
PR 19-APR-1999; 99US-0130077.
PR 21-APR-1999; 99US-0130449.
PR 23-APR-1999; 99US-0130510.
PR 23-APR-1999; 99US-0130891.
PR 28-APR-1999; 99US-0131449.
PR 30-APR-1999; 99US-0132048.
PR 30-APR-1999; 99US-0132407.
PR 04-MAY-1999; 99US-0132484.
PR 05-MAY-1999; 99US-0132485.
PR 06-MAY-1999; 99US-0132486.
PR 06-MAY-1999; 99US-0132487.
PR 07-MAY-1999; 99US-0132863.
PR 14-MAY-1999; 99US-0134218.
PR 14-MAY-1999; 99US-0134219.
PR 14-MAY-1999; 99US-0134221.
PR 14-MAY-1999; 99US-0134370.
PR 18-MAY-1999; 99US-0134768.
PR 19-MAY-1999; 99US-0134941.
PR 20-MAY-1999; 99US-0135124.
PR 21-MAY-1999; 99US-0135353.
PR 25-MAY-1999; 99US-0135629.
PR 25-MAY-1999; 99US-0136021.
PR 27-MAY-1999; 99US-0136382.
PR 28-MAY-1999; 99US-0136782.
PR 01-JUN-1999; 99US-0137222.
PR 03-JUN-1999; 99US-0137528.
PR 04-JUN-1999; 99US-0137502.
PR 07-JUN-1999; 99US-0137724.
PR 08-JUN-1999; 99US-0138094.
PR 10-JUN-1999; 99US-0138540.
PR 10-JUN-1999; 99US-0138847.
PR 14-JUN-1999; 99US-0139119.
PR 16-JUN-1999; 99US-0139452.
PR 16-JUN-1999; 99US-0139453.
PR 17-JUN-1999; 99US-0139452.
PR 18-JUN-1999; 99US-0139454.
PR 18-JUN-1999; 99US-0139455.
PR 18-JUN-1999; 99US-0139456.
PR 18-JUN-1999; 99US-0139457.
PR 18-JUN-1999; 99US-0139458.
PR 18-JUN-1999; 99US-0139459.
PR 18-JUN-1999; 99US-0139460.
PR 18-JUN-1999; 99US-0139461.
PR 18-JUN-1999; 99US-0139462.
PR 18-JUN-1999; 99US-0139463.
PR 18-JUN-1999; 99US-0139750.
PR 18-JUN-1999; 99US-0139753.
PR 21-JUN-1999; 99US-0139817.
PR 22-JUN-1999; 99US-0139899.
PR 23-JUN-1999; 99US-0140353.
PR 23-JUN-1999; 99US-0140354.
PR 24-JUN-1999; 99US-0140695.
PR 28-JUN-1999; 99US-0140823.
PR 29-JUN-1999; 99US-0140991.
PR 30-JUN-1999; 99US-0141287.
PR 01-JUL-1999; 99US-0141842.
PR 01-JUL-1999; 99US-0142154.
PR 02-JUL-1999; 99US-0142055.
PR 06-JUL-1999; 99US-0142390.

OY 491 DQPAFFPVPDNGLPFRVSR 511
| : | | | |
Db 466 DKLVFFPTTRQKRYPIFVKR 486

RESULT 6
AAG4570
ID AAG4570 standard; Protein: 492 AA.
XX AAG4570;
AC
XX 18-OCT-2000 (first entry)
DT
XX Arabidopsis thaliana protein fragment SEQ ID NO: 55846.
DE
XX Protein identification: signal transduction pathway; metabolic pathway;
KW hybridisation assay; genetic mapping; gene expression control; promoter;
KW termination sequence.
XX Arabidopsis thaliana.
XX EP1033405-A2.
PN
XX 06-SEP-2000.
PD
XX 25-FEB-2000; 2000EP-0301439.
PE
XX 25-FEB-1999; 99US-0121825.
XX 05-MAR-1999; 99US-0123180.
PR 09-MAR-1999; 99US-0123548.
PR 23-MAR-1999; 99US-0125788.
PR 23-MAR-1999; 99US-0126264.
PR 29-MAR-1999; 99US-0126785.
PR 01-APR-1999; 99US-0127462.
PR 06-APR-1999; 99US-0128234.
PR 08-APR-1999; 99US-0128714.
PR 16-APR-1999; 99US-0129845.
PR 19-APR-1999; 99US-0130077.
PR 21-APR-1999; 99US-0130449.
PR 23-APR-1999; 99US-0130510.
PR 28-APR-1999; 99US-0130891.
PR 30-APR-1999; 99US-0131449.
PR 30-APR-1999; 99US-0132048.
PR 30-APR-1999; 99US-0132407.
PR 04-MAY-1999; 99US-0132484.
PR 05-MAY-1999; 99US-0132485.
PR 06-MAY-1999; 99US-0132486.
PR 07-MAY-1999; 99US-0132487.
PR 11-MAY-1999; 99US-0132863.
PR 11-MAY-1999; 99US-0134256.
PR 14-MAY-1999; 99US-0134218.
PR 14-MAY-1999; 99US-0134219.
PR 14-MAY-1999; 99US-0134221.
PR 14-MAY-1999; 99US-0134370.
PR 18-MAY-1999; 99US-0134768.
PR 19-MAY-1999; 99US-0134941.
PR 20-MAY-1999; 99US-0135124.
PR 21-MAY-1999; 99US-0135353.
PR 24-MAY-1999; 99US-0135629.
PR 25-MAY-1999; 99US-0136021.
PR 27-MAY-1999; 99US-0136382.
PR 28-MAY-1999; 99US-0136782.
PR 01-JUN-1999; 99US-0137222.
PR 03-JUN-1999; 99US-0137528.
PR 04-JUN-1999; 99US-0137502.
PR 07-JUN-1999; 99US-0137724.
PR 08-JUN-1999; 99US-0138094.
PR 10-JUN-1999; 99US-0138540.
PR 10-JUN-1999; 99US-0138847.
PR 14-JUN-1999; 99US-0139119.
PR 16-JUN-1999; 99US-0139452.
PR 16-JUN-1999; 99US-0139453.

PR 17-JUN-1999; 99US-0139492.
PR 18-JUN-1999; 99US-0139454.
PR 18-JUN-1999; 99US-0139455.
PR 18-JUN-1999; 99US-0139456.
PR 18-JUN-1999; 99US-0139457.
PR 18-JUN-1999; 99US-0139458.
PR 18-JUN-1999; 99US-0139459.
PR 18-JUN-1999; 99US-0139460.
PR 18-JUN-1999; 99US-0139461.
PR 18-JUN-1999; 99US-0139462.
PR 18-JUN-1999; 99US-0139463.
PR 18-JUN-1999; 99US-0139750.
PR 18-JUN-1999; 99US-0139763.
PR 21-JUN-1999; 99US-0139817.
PR 22-JUN-1999; 99US-0139899.
PR 23-JUN-1999; 99US-0140353.
PR 23-JUN-1999; 99US-0140354.
PR 24-JUN-1999; 99US-0140695.
PR 28-JUN-1999; 99US-0140823.
PR 29-JUN-1999; 99US-0140991.
PR 30-JUN-1999; 99US-0141287.
PR 01-JUL-1999; 99US-0141842.
PR 01-JUL-1999; 99US-0142154.
PR 02-JUL-1999; 99US-0142055.
PR 06-JUL-1999; 99US-0142390.
PR 08-JUL-1999; 99US-0142803.
PR 09-JUL-1999; 99US-0142920.
PR 12-JUL-1999; 99US-0142977.
PR 13-JUL-1999; 99US-0143542.
PR 14-JUL-1999; 99US-0143624.
PR 15-JUL-1999; 99US-0144005.
PR 16-JUL-1999; 99US-0144085.
PR 16-JUL-1999; 99US-0144086.
PR 19-JUL-1999; 99US-0144325.
PR 19-JUL-1999; 99US-0144331.
PR 19-JUL-1999; 99US-0144332.
PR 19-JUL-1999; 99US-0144333.
PR 19-JUL-1999; 99US-0144334.
PR 20-JUL-1999; 99US-0144335.
PR 20-JUL-1999; 99US-0144352.
PR 20-JUL-1999; 99US-0144632.
PR 20-JUL-1999; 99US-0144884.
PR 21-JUL-1999; 99US-0144814.
PR 21-JUL-1999; 99US-0145086.
PR 21-JUL-1999; 99US-0145086.
PR 22-JUL-1999; 99US-0145085.
PR 22-JUL-1999; 99US-0145087.
PR 22-JUL-1999; 99US-0145089.
PR 22-JUL-1999; 99US-0145192.
PR 23-JUL-1999; 99US-0145145.
PR 23-JUL-1999; 99US-0145218.
PR 23-JUL-1999; 99US-0145224.
PR 26-JUL-1999; 99US-0145276.
PR 27-JUL-1999; 99US-0145913.
PR 27-JUL-1999; 99US-0145918.
PR 27-JUL-1999; 99US-0145919.
PR 28-JUL-1999; 99US-0145951.
PR 02-AUG-1999; 99US-0146386.
PR 02-AUG-1999; 99US-0146388.
PR 02-AUG-1999; 99US-0146389.
PR 03-AUG-1999; 99US-0147038.
PR 04-AUG-1999; 99US-0147204.
PR 04-AUG-1999; 99US-0147302.
PR 05-AUG-1999; 99US-0147260.
PR 06-AUG-1999; 99US-0147303.
PR 06-AUG-1999; 99US-0147416.
PR 09-AUG-1999; 99US-0147493.
PR 10-AUG-1999; 99US-0147935.
PR 11-AUG-1999; 99US-0148171.
PR 12-AUG-1999; 99US-0148319.
PR 13-AUG-1999; 99US-0148565.

```

PR 13-AUG-1999; 990S-0148684.
PR 16-AUG-1999; 990S-0149368.
PR 17-AUG-1999; 990S-0149175.
PR 18-AUG-1999; 990S-0149426.
PR 20-AUG-1999; 990S-0149722.
PR 20-AUG-1999; 990S-0149722.
PR 20-AUG-1999; 990S-0149722.
PR 23-AUG-1999; 990S-0149929.
PR 23-AUG-1999; 990S-0149902.
PR 25-AUG-1999; 990S-0149930.
PR 26-AUG-1999; 990S-0150566.
PR 26-AUG-1999; 990S-0150884.
PR 27-AUG-1999; 990S-0151065.
PR 27-AUG-1999; 990S-0151066.
PR 30-AUG-1999; 990S-0151080.
PR 31-AUG-1999; 990S-0151438.
PR 01-SEP-1999; 990S-0151930.
PR 07-SEP-1999; 990S-0152363.
PR 10-SEP-1999; 990S-0153070.
PR 13-SEP-1999; 990S-0153758.
PR 15-SEP-1999; 990S-0154018.
PR 16-SEP-1999; 990S-0154039.
PR 20-SEP-1999; 990S-0154779.
PR 22-SEP-1999; 990S-0155139.
PR 23-SEP-1999; 990S-0155486.
PR 24-SEP-1999; 990S-0155659.
PR 28-SEP-1999; 990S-0156458.
PR 29-SEP-1999; 990S-0156596.
PR 04-OCT-1999; 990S-0157117.
PR 05-OCT-1999; 990S-0157753.
PR 06-OCT-1999; 990S-0157865.
PR 07-OCT-1999; 990S-0158029.
PR 08-OCT-1999; 990S-0158232.
PR 12-OCT-1999; 990S-0158369.
PR 13-OCT-1999; 990S-0159293.
PR 13-OCT-1999; 990S-0159294.
PR 14-OCT-1999; 990S-0159295.
PR 14-OCT-1999; 990S-0159329.
PR 14-OCT-1999; 990S-0159330.
PR 14-OCT-1999; 990S-0159331.
PR 14-OCT-1999; 990S-0159637.
PR 18-OCT-1999; 990S-0159638.
PR 21-OCT-1999; 990S-0159584.
PR 21-OCT-1999; 990S-0160741.
PR 21-OCT-1999; 990S-0160767.
PR 21-OCT-1999; 990S-0160768.
PR 21-OCT-1999; 990S-0160770.
PR 21-OCT-1999; 990S-0160814.
PR 21-OCT-1999; 990S-0160815.
PR 22-OCT-1999; 990S-0160980.
PR 22-OCT-1999; 990S-0160981.
PR 22-OCT-1999; 990S-0160989.
PR 25-OCT-1999; 990S-0161404.
PR 25-OCT-1999; 990S-0161405.
PR 25-OCT-1999; 990S-0161406.
PR 26-OCT-1999; 990S-0161359.
PR 26-OCT-1999; 990S-0161360.
PR 26-OCT-1999; 990S-0161361.
PR 28-OCT-1999; 990S-0161920.
PR 28-OCT-1999; 990S-0161922.
PR 28-OCT-1999; 990S-0161993.
PR 29-OCT-1999; 990S-0162142.

```

Query Match 36.5%; Score 978.5; DB 21; Length 492;
 Best Local Similarity 41.9%; Pred. No. 7e-78;
 Matches 210; Conservative 79; Mismatches 171; Indels 41; Gaps 8;

```

QY 12 LLLPSSLSLLFTILKRRNRKTRFNLPGKSGMPFLGETIGYIKPYTATTLGDFMOQH 71
DB 27 LLLSSIAAGFL--LLLRTRRRMGGLPGSLPLIGERFOLIGAKTKENPEPFIDER 83
QY 72 VSKYGIKIRSNLFEPTIVSDAGLNFRIIONERKLEECSTPRISGILGRKMSMLVYGD 131
|::||:::||||||| ||| ||:||||:||||||| || |||||:::|

```

```

DB 84 VARYGSVMTHLEGEPTIFSADEPTNRFVLONEGKLEECYPASICNLGKSHLLMKGS 143
QY 132 MHRMRSISLNFSLHALKRLTILKQVBRHTLFLYIDSMOONSIFSADEAKKFTFNIMARH 191
DB 144 LHKMHSLLTMSFANSSITIKDMLDIDRLVRFNIDSMSSRVLL--MEAKKIFELTVQ 201
QY 192 IMSMDPEEETEDLKKRYVFMKGVSAPLNLGCTAVHKAOSRATILNFIKKMERKL 251
DB 202 IMSFDRG-EWSESLRKRYLLVIGFSLPLPLSTTYKRAIQAR-----RKVAELT 252
QY 252 DIKEEDQEEEVTEDEAEKSKSDHYRKORTDDLLGWYLKHSNLSTEQILDLISLPLA 311
DB 253 VVVKRRREEEGAE-----RKMDLAALLAADGFEDEIVPIVALVA 298
QY 312 GHETSSVAIALAIFLOACPKAVEELREHELEIARAKKLGSELMWDYKKMDFTOCY 371
DB 299 GYETITSTIMLANKFLTEPLALQLEHEHEKIRANKSD--SYLSWSDYKSMPTOCYV 356
QY 372 NETLRIGNVYRFLHRRALDVRYKGYDIPSGKVLPIVSAVHLDSRYDQPLFNPWRMO 431
DB 357 NETLRVANIILIGVFRRAMTDVEIKGYKIPKMKVSSFFRAVILDRNHRDANTEFNPWRMO 416
QY 432 QQNGASSSGSGSFTWGN--YMPRGCGPRLCAGSELALEMAVFTHLVLFKNWLEAD 490
DB 417 -----SNSVTTGPSNVFTPGGGPRLCPGELARVALSVFLHRLVTGFSWVPAEQ 466
QY 491 DQFAFPFVDFPNGPLRIVRSR 511
DB 467 DLVFEPTTRTQKRRYPPIYKR 487

RESULT 7
AA644572
ID AA64572 standard; Protein; 444 AA.
XX
AC AA64572;
XX
DT 18-OCT-2000 (first entry)
XX
DE Arabidopsis thaliana protein fragment SEQ ID NO: 55848.
XX
KW Protein identification; signal transduction pathway; metabolic pathway;
KW hybridisation assay; genetic mapping; gene expression control; promoter;
KW termination sequence.
XX
OS Arabidopsis thaliana.
XX
PN EP1033405-A2.
XX
PD 06-SEP-2000.
XX
PE 25-FEB-2000; 2000EP-0301439.
XX
PR 25-FEB-1999; 990S-0121825.
PR 05-MAR-1999; 990S-0123180.
PR 09-MAR-1999; 990S-0123548.
PR 23-MAR-1999; 990S-0125784.
PR 25-MAR-1999; 990S-0126284.
PR 29-MAR-1999; 990S-0126785.
PR 01-APR-1999; 990S-0127462.
PR 06-APR-1999; 990S-0128234.
PR 08-APR-1999; 990S-0128714.
PR 16-APR-1999; 990S-0129845.
PR 19-APR-1999; 990S-0130077.
PR 21-APR-1999; 990S-0130449.
PR 23-APR-1999; 990S-0130510.
PR 28-APR-1999; 990S-0130891.
PR 30-APR-1999; 990S-0131449.
PR 30-APR-1999; 990S-0132048.
PR 04-MAY-1999; 990S-0132407.
PR 05-MAY-1999; 990S-0132484.
PR 06-MAY-1999; 990S-0132485.
PR 990S-0132486.

```

PR 06-MAY-1999; 99US-0132487.
PR 07-MAY-1999; 99US-0132863.
PR 11-MAY-1999; 99US-0134256.
PR 14-MAY-1999; 99US-0134218.
PR 14-MAY-1999; 99US-0134219.
PR 14-MAY-1999; 99US-0134221.
PR 14-MAY-1999; 99US-0134370.
PR 14-MAY-1999; 99US-0134768.
PR 19-MAY-1999; 99US-0135124.
PR 20-MAY-1999; 99US-0135124.
PR 21-MAY-1999; 99US-0135353.
PR 24-MAY-1999; 99US-0135629.
PR 25-MAY-1999; 99US-0136021.
PR 27-MAY-1999; 99US-0136382.
PR 28-MAY-1999; 99US-0136782.
PR 01-JUN-1999; 99US-0137222.
PR 03-JUN-1999; 99US-0137528.
PR 04-JUN-1999; 99US-0137502.
PR 07-JUN-1999; 99US-0137724.
PR 08-JUN-1999; 99US-0138094.
PR 10-JUN-1999; 99US-0138540.
PR 10-JUN-1999; 99US-0138847.
PR 14-JUN-1999; 99US-0139119.
PR 16-JUN-1999; 99US-0139452.
PR 17-JUN-1999; 99US-0139453.
PR 17-JUN-1999; 99US-0139492.
PR 18-JUN-1999; 99US-0139454.
PR 18-JUN-1999; 99US-0139455.
PR 18-JUN-1999; 99US-0139456.
PR 18-JUN-1999; 99US-0139457.
PR 18-JUN-1999; 99US-0139458.
PR 18-JUN-1999; 99US-0139459.
PR 18-JUN-1999; 99US-0139460.
PR 18-JUN-1999; 99US-0139461.
PR 18-JUN-1999; 99US-0139462.
PR 18-JUN-1999; 99US-0139463.
PR 18-JUN-1999; 99US-0139750.
PR 18-JUN-1999; 99US-0139763.
PR 21-JUN-1999; 99US-0139817.
PR 22-JUN-1999; 99US-0139899.
PR 23-JUN-1999; 99US-0140353.
PR 23-JUN-1999; 99US-0140354.
PR 24-JUN-1999; 99US-0140685.
PR 28-JUN-1999; 99US-0140823.
PR 29-JUN-1999; 99US-0140991.
PR 30-JUN-1999; 99US-0141287.
PR 01-JUL-1999; 99US-0141842.
PR 01-JUL-1999; 99US-0142154.
PR 02-JUL-1999; 99US-0142055.
PR 06-JUL-1999; 99US-0142380.
PR 08-JUL-1999; 99US-0142803.
PR 09-JUL-1999; 99US-0142920.
PR 12-JUL-1999; 99US-0142977.
PR 13-JUL-1999; 99US-0143542.
PR 14-JUL-1999; 99US-0143624.
PR 15-JUL-1999; 99US-0144005.
PR 16-JUL-1999; 99US-0144085.
PR 16-JUL-1999; 99US-0144086.
PR 19-JUL-1999; 99US-0144325.
PR 19-JUL-1999; 99US-0144331.
PR 19-JUL-1999; 99US-0144332.
PR 19-JUL-1999; 99US-0144333.
PR 19-JUL-1999; 99US-0144334.
PR 19-JUL-1999; 99US-0144335.
PR 20-JUL-1999; 99US-0144352.
PR 20-JUL-1999; 99US-0144632.
PR 20-JUL-1999; 99US-0144634.
PR 21-JUL-1999; 99US-0144814.
PR 21-JUL-1999; 99US-0145086.
PR 21-JUL-1999; 99US-0145088.
PR 22-JUL-1999; 99US-0145085.
PR 22-JUL-1999; 99US-0145087.
PR 22-JUL-1999; 99US-0145089.

PR 22-JUL-1999; 99US-0145192.
PR 23-JUL-1999; 99US-0145145.
PR 23-JUL-1999; 99US-0145218.
PR 23-JUL-1999; 99US-0145224.
PR 26-JUL-1999; 99US-0145276.
PR 27-JUL-1999; 99US-0145913.
PR 27-JUL-1999; 99US-0145919.
PR 27-JUL-1999; 99US-0145919.
PR 28-JUL-1999; 99US-0145919.
PR 28-JUL-1999; 99US-0145919.
PR 02-AUG-1999; 99US-0146388.
PR 02-AUG-1999; 99US-0146388.
PR 03-AUG-1999; 99US-0146388.
PR 04-AUG-1999; 99US-0147204.
PR 04-AUG-1999; 99US-0147204.
PR 05-AUG-1999; 99US-0147192.
PR 05-AUG-1999; 99US-0147260.
PR 06-AUG-1999; 99US-0147302.
PR 06-AUG-1999; 99US-0147418.
PR 09-AUG-1999; 99US-0147493.
PR 09-AUG-1999; 99US-0147935.
PR 10-AUG-1999; 99US-0148171.
PR 11-AUG-1999; 99US-0148319.
PR 12-AUG-1999; 99US-0148341.
PR 13-AUG-1999; 99US-0148365.
PR 13-AUG-1999; 99US-0148684.
PR 16-AUG-1999; 99US-0149368.
PR 17-AUG-1999; 99US-0149175.
PR 18-AUG-1999; 99US-0149426.
PR 20-AUG-1999; 99US-0149722.
PR 20-AUG-1999; 99US-0149722.
PR 20-AUG-1999; 99US-0149723.
PR 20-AUG-1999; 99US-0149929.
PR 23-AUG-1999; 99US-0149902.
PR 23-AUG-1999; 99US-0149930.
PR 25-AUG-1999; 99US-0150566.
PR 26-AUG-1999; 99US-0150884.
PR 27-AUG-1999; 99US-0151066.
PR 27-AUG-1999; 99US-0151066.
PR 27-AUG-1999; 99US-0151086.
PR 30-AUG-1999; 99US-0151303.
PR 31-AUG-1999; 99US-0151438.
PR 01-SEP-1999; 99US-0151930.
PR 07-SEP-1999; 99US-0152363.
PR 10-SEP-1999; 99US-0153070.
PR 13-SEP-1999; 99US-0153758.
PR 15-SEP-1999; 99US-0154018.
PR 16-SEP-1999; 99US-0154039.
PR 20-SEP-1999; 99US-0154779.
PR 22-SEP-1999; 99US-0155139.
PR 23-SEP-1999; 99US-0155486.
PR 24-SEP-1999; 99US-0155659.
PR 28-SEP-1999; 99US-0156458.
PR 29-SEP-1999; 99US-0156596.
PR 04-OCT-1999; 99US-0157117.
PR 05-OCT-1999; 99US-0157753.
PR 06-OCT-1999; 99US-0157865.
PR 07-OCT-1999; 99US-0158029.
PR 08-OCT-1999; 99US-0158232.
PR 12-OCT-1999; 99US-0158369.
PR 13-OCT-1999; 99US-0159293.
PR 13-OCT-1999; 99US-0159294.
PR 13-OCT-1999; 99US-0159295.
PR 14-OCT-1999; 99US-0159329.
PR 14-OCT-1999; 99US-0159330.
PR 14-OCT-1999; 99US-0159331.
PR 14-OCT-1999; 99US-0159637.
PR 14-OCT-1999; 99US-0159638.
PR 18-OCT-1999; 99US-0159584.
PR 21-OCT-1999; 99US-0160741.
PR 21-OCT-1999; 99US-0160767.
PR 21-OCT-1999; 99US-0160768.
PR 21-OCT-1999; 99US-0160770.
PR 21-OCT-1999; 99US-0160814.

XX	25-FEB-1999,	9905-0121825Z
PR	05-MAR-1999,	9905-01231180
PR	09-MAR-1999,	9905-01235449
PR	23-MAR-1999,	9905-01257888
PR	25-MAR-1999,	9905-01262666
PR	29-MAR-1999,	9905-01267595
PR	01-APR-1999,	9905-01274652
PR	06-APR-1999,	9905-01288274
PR	08-APR-1999,	9905-01287244
PR	16-APR-1999,	9905-01298455
PR	19-APR-1999,	9905-01300777
PR	21-APR-1999,	9905-01300499
PR	23-APR-1999,	9905-01305010
PR	23-APR-1999,	9905-01308911
PR	28-APR-1999,	9905-01314449
PR	30-APR-1999,	9905-01320408
PR	04-MAY-1999,	9905-01324078
PR	04-MAY-1999,	9905-01324844
PR	05-MAY-1999,	9905-01324853
PR	06-MAY-1999,	9905-01324866
PR	07-MAY-1999,	9905-01324877
PR	11-MAY-1999,	9905-01328653
PR	14-MAY-1999,	9905-01342118
PR	14-MAY-1999,	9905-01342129
PR	14-MAY-1999,	9905-01343271
PR	14-MAY-1999,	9905-01343701
PR	18-MAY-1999,	9905-01347680
PR	19-MAY-1999,	9905-01349141
PR	21-MAY-1999,	9905-01351414
PR	21-MAY-1999,	9905-01353533
PR	25-MAY-1999,	9905-01356299
PR	27-MAY-1999,	9905-01360711
PR	28-MAY-1999,	9905-01363632
PR	01-JUN-1999,	9905-01367682
PR	03-JUN-1999,	9905-01372528
PR	04-JUN-1999,	9905-01375702
PR	07-JUN-1999,	9905-01377244
PR	08-JUN-1999,	9905-01380604
PR	10-JUN-1999,	9905-01385400
PR	14-JUN-1999,	9905-01388477
PR	16-JUN-1999,	9905-01391412
PR	18-JUN-1999,	9905-01394532
PR	18-JUN-1999,	9905-01394545
PR	18-JUN-1999,	9905-01394655
PR	18-JUN-1999,	9905-01394673
PR	18-JUN-1999,	9905-01397650
PR	21-JUN-1999,	9905-01396177
PR	22-JUN-1999,	9905-01396699
PR	23-JUN-1999,	9905-01403553
PR	24-JUN-1999,	9905-01403554
PR	28-JUN-1999,	9905-01406095
PR	29-JUN-1999,	9905-01409823
PR	30-JUN-1999,	9905-01409931
PR	01-JUL-1999,	9905-01412877
PR	01-JUL-1999,	9905-01414842
PR	02-JUL-1999,	9905-01421554
PR	06-JUL-1999,	9905-01423290
PR	08-JUL-1999,	9905-01428203
PR	09-JUL-1999,	9905-01429200

PR 12-JUL-1999; 99US-0142977.
PR 13-JUL-1999; 99US-0143542.
PR 14-JUL-1999; 99US-0143624.
PR 15-JUL-1999; 99US-0144005.
PR 16-JUL-1999; 99US-0144085.
PR 16-JUL-1999; 99US-0144086.
PR 19-JUL-1999; 99US-0144325.
PR 19-JUL-1999; 99US-0144331.
PR 19-JUL-1999; 99US-0144332.
PR 19-JUL-1999; 99US-0144333.
PR 19-JUL-1999; 99US-0144334.
PR 19-JUL-1999; 99US-0144335.
PR 20-JUL-1999; 99US-0144352.
PR 20-JUL-1999; 99US-0144632.
PR 20-JUL-1999; 99US-0144864.
PR 21-JUL-1999; 99US-0145086.
PR 21-JUL-1999; 99US-0145087.
PR 22-JUL-1999; 99US-0145088.
PR 22-JUL-1999; 99US-0145087.
PR 22-JUL-1999; 99US-0145089.
PR 22-JUL-1999; 99US-0145192.
PR 23-JUL-1999; 99US-0145145.
PR 23-JUL-1999; 99US-0145218.
PR 23-JUL-1999; 99US-0145224.
PR 26-JUL-1999; 99US-0145276.
PR 27-JUL-1999; 99US-0145913.
PR 27-JUL-1999; 99US-0145918.
PR 27-JUL-1999; 99US-0145919.
PR 28-JUL-1999; 99US-0145951.
PR 02-AUG-1999; 99US-0146386.
PR 02-AUG-1999; 99US-0146388.
PR 02-AUG-1999; 99US-0146389.
PR 03-AUG-1999; 99US-0147038.
PR 04-AUG-1999; 99US-0147204.
PR 04-AUG-1999; 99US-0147302.
PR 05-AUG-1999; 99US-0147312.
PR 05-AUG-1999; 99US-0147260.
PR 06-AUG-1999; 99US-0147303.
PR 06-AUG-1999; 99US-0147416.
PR 09-AUG-1999; 99US-0147493.
PR 10-AUG-1999; 99US-0147935.
PR 11-AUG-1999; 99US-0148171.
PR 11-AUG-1999; 99US-0148319.
PR 12-AUG-1999; 99US-0148341.
PR 13-AUG-1999; 99US-0148565.
PR 13-AUG-1999; 99US-0148684.
PR 16-AUG-1999; 99US-0149368.
PR 17-AUG-1999; 99US-0149175.
PR 18-AUG-1999; 99US-0149426.
PR 20-AUG-1999; 99US-0149722.
PR 20-AUG-1999; 99US-0149723.
PR 20-AUG-1999; 99US-0149929.
PR 23-AUG-1999; 99US-0149902.
PR 23-AUG-1999; 99US-0149930.
PR 25-AUG-1999; 99US-0150566.
PR 26-AUG-1999; 99US-0150884.
PR 27-AUG-1999; 99US-0151065.
PR 27-AUG-1999; 99US-0151066.
PR 27-AUG-1999; 99US-0151080.
PR 30-AUG-1999; 99US-0151303.
PR 31-AUG-1999; 99US-0151438.
PR 01-SEP-1999; 99US-0151930.
PR 07-SEP-1999; 99US-0152363.
PR 10-SEP-1999; 99US-0153070.
PR 13-SEP-1999; 99US-0153758.
PR 15-SEP-1999; 99US-0154018.
PR 16-SEP-1999; 99US-0154039.
PR 20-SEP-1999; 99US-0154779.
PR 22-SEP-1999; 99US-0155139.
PR 23-SEP-1999; 99US-0155486.
PR 24-SEP-1999; 99US-0155659.
PR 28-SEP-1999; 99US-0156458.

PR 29-SEP-1999; 99US-0156596.
PR 04-OCT-1999; 99US-0157117.
PR 05-OCT-1999; 99US-0157753.
PR 06-OCT-1999; 99US-0157865.
PR 07-OCT-1999; 99US-0158029.
PR 08-OCT-1999; 99US-0158232.
PR 12-OCT-1999; 99US-0158366.
PR 13-OCT-1999; 99US-0159293.
PR 13-OCT-1999; 99US-0159294.
PR 13-OCT-1999; 99US-0159295.
PR 14-OCT-1999; 99US-0159320.
PR 14-OCT-1999; 99US-0159330.
PR 14-OCT-1999; 99US-0159331.
PR 14-OCT-1999; 99US-0159637.
PR 14-OCT-1999; 99US-0159638.
PR 18-OCT-1999; 99US-0159584.
PR 21-OCT-1999; 99US-0160741.
PR 21-OCT-1999; 99US-0160767.
PR 21-OCT-1999; 99US-0160770.
PR 21-OCT-1999; 99US-0160770.
PR 21-OCT-1999; 99US-0160814.
PR 21-OCT-1999; 99US-0160815.
PR 22-OCT-1999; 99US-0160980.
PR 22-OCT-1999; 99US-0160981.
PR 22-OCT-1999; 99US-0160989.
PR 25-OCT-1999; 99US-0161405.
PR 25-OCT-1999; 99US-0161405.
PR 25-OCT-1999; 99US-0161406.
PR 26-OCT-1999; 99US-0161359.
PR 26-OCT-1999; 99US-0161360.
PR 26-OCT-1999; 99US-0161361.
PR 28-OCT-1999; 99US-0161920.
PR 28-OCT-1999; 99US-0161992.
PR 28-OCT-1999; 99US-0161993.
PR 29-OCT-1999; 99US-0162142.

Query Match 35.3%; Score 946; DB 21; Length 444;
Best Local Similarity 41.8%; Pred. No. 4.6e-75;
Matches 198; Conservative 77; Mismatches 161; Indels 38; Gaps 7;

QY 39 LPFGSGWPLGELIGVTKPTATTLDGFMQOHKSKYKIRSNLFGSEPTVSADAGNLR 98
DB 3 LPFGSLGRLGELIGVTKPTATTLDGFMQOHKSKYKIRSNLFGSEPTVSADAGNLR 62
QY 99 FTLNKGRLFECSYPRISGILGKWSMLVLDGMDHRSISLNFSLHARLTILKDYE 158
DB 63 FVLNKGRLFECSYPRISGILGKWSMLVLDGMDHRSISLNFSLHARLTILKDYE 122
QY 159 RHTLFVLDSDQOINIFSAODEAKFTFNLMKHHMSMDPGEETDQKKEVTFMKGVYS 218
DB 123 RLVRFNLDSSWSRYLL--MEBAKRTFELTVKQLMSFDPG--EWSLSLKEVLYEGFFS 179
QY 219 APNLPGAYKKAALOSRATILKFTERKKEERKLDIKEDDOEEVKTDEAEMSDHVR 278
DB 180 LPFLFSTYTKALQAR-----KVAELTVVVMKRREEBGAE----- 219
QY 279 KQRTDDLLGVNLKSHNSLSTEQIIDLILSLFAGHETSVAIALIFELQACPAVEELR 338
DB 220 --RKMDLMAALLAADGFSDEIYDPLVALLVAGYETSTIMTLAVKFLTEPLALQGLK 277
QY 339 EEHLFIARAKKEGESELMNDYKMDQTCVINEYETLRLGVNVRFLHRAKALDKVRYKCYD 398
DB 278 EEHEKIRAMKSD--SYSLSEMSDYKSMPTQCVNETLRVANIGGVFRAMTDVEIKGYK 335
QY 399 IPGSKVLPYSAVYHLSNSRYDQNLNFPNMQOONNGASSGSGSFTMGN--YMPRGK 457
DB 336 IPGSKVLPYSAVYHLSNSRYDQNLNFPNMQOONNGASSGSGSFTMGN--YMPRGK 385
QY 458 GPRLCAGSELAKEMAVFIHHLVLYKFMWELAEDDQPAFPVDPNGLPIRVSR 511
DB 386 GPRLCAGSELAKEMAVFIHHLVLYKFMWELAEDDQPAFPVDPNGLPIRVSR 439

QY 73 SKGKTYRSNLFGEPR-ITVSDADLNRFITLONEGRLEBCSPRSIGGILGWSMLVYGD 131
Db 80 LRYGPITFRCSLFGSTRVAVSDPPEFNKYVLONEGRLEFSNALAFERNLIGKYSAYGGE 139
QY 132 MHRMRSISLNFELSHARLRTLLKDVREHRTLFVLDMSQOONSIFSAODAKKFTFNMAKH 191
Db 140 LQRLHATVAVLLHHELLSSDFMEDIDIFQAGKMKKEEGDIPLOHKCNOIVLNLAKR 199
QY 192 IMSNDPGEETEOLKKEVTFPMKGVSAPLNPCTAVYHKAQSRATILKFEYRKMEEKRL 251
Db 200 LLDLP-SEEMGHLYKAFDDPVGAVLSPLNIPCTYARGARAILLKRHKCKERR- 257
QY 252 DIKEEDDEBEVEVTKEDAEKMSDVRKQRTDDLLGVNLKHSMLSTEQIIDLISLLFA 311
Db 258 -----EHPEV-----LRNDLTKLVREGTSEDEIADTIIFVEFA 292
QY 312 GHETSSVAIALAIFELQACPAVVELREHELEIRAKKEGSESLNMDYKKNMTOCVI 371
Db 293 GVEISAMAMTFVAVYLLDENPRALELRAHEDALAKAGK-GENEKLTMNDYQSMKFEVHCVI 351
QY 372 NETLRIGNVVFLHRAKALDKVRYKGYDIPSGMKVLPVISAHLNDSRYDQNLFPNPMQ 431
Db 352 NETLRIGGATVFLERAKKODIKVDVLPKGMTVSVELSATHVDGKHYEADKFLPMRMQ 411
QY 432 QQNGGASSGSGSFSTGNNTMPGCGPRLCAGSELAKLEMAVTHHLVLFKNMELADD 491
Db 412 NE-----GQETLEPCYMPRGGRGLCPGLHARFETALFLHNFVTFRMEQLEID 462
QY 492 QPFAFPVDFPNGLPIRV-SRI 512
Db 463 RAYTFPLPSTENGPRILYSRV 484

RESULT 11
AAG46490
ID AAG46490 standard; Protein; 461 AA.
XX
AC AAG46490;
XX
DT 18-OCT-2000 (first entry)
XX
DE Arabidopsis thaliana protein fragment SEQ ID NO: 58494.
XX
KW Protein identification; signal transduction pathway; metabolic pathway;
KW hybridisation assay; genetic mapping; gene expression control; promoter;
XX
XX
OS Arabidopsis thaliana.
XX
PN EP1033405-A2.
XX
PD 06-SEP-2000.
XX
PF 25-FEB-2000; 2000EP-0301439.
XX
PR 25-FEB-1999; 99US-0121825.
PR 03-MAR-1999; 99US-0123180.
PR 09-MAR-1999; 99US-0123548.
PR 23-MAR-1999; 99US-0125788.
PR 25-MAR-1999; 99US-0126264.
PR 29-MAR-1999; 99US-0126785.
PR 01-APR-1999; 99US-0127462.
PR 06-APR-1999; 99US-0128234.
PR 08-APR-1999; 99US-0128714.
PR 16-APR-1999; 99US-0129845.
PR 19-APR-1999; 99US-0130077.
PR 21-APR-1999; 99US-0130449.
PR 23-APR-1999; 99US-0130510.
PR 23-APR-1999; 99US-0130891.
PR 28-APR-1999; 99US-0131449.
PR 30-APR-1999; 99US-0132048.
PR 30-APR-1999; 99US-0132407.

PR 04-MAY-1999; 99US-0132484.
PR 05-MAY-1999; 99US-0132485.
PR 06-MAY-1999; 99US-0132486.
PR 07-MAY-1999; 99US-0132487.
PR 11-MAY-1999; 99US-0132488.
PR 14-MAY-1999; 99US-0132489.
PR 14-MAY-1999; 99US-0134219.
PR 14-MAY-1999; 99US-0134221.
PR 14-MAY-1999; 99US-0134370.
PR 18-MAY-1999; 99US-0134768.
PR 19-MAY-1999; 99US-0134941.
PR 20-MAY-1999; 99US-0135124.
PR 21-MAY-1999; 99US-0135353.
PR 24-MAY-1999; 99US-0135629.
PR 25-MAY-1999; 99US-0136021.
PR 27-MAY-1999; 99US-0136392.
PR 28-MAY-1999; 99US-0136782.
PR 01-JUN-1999; 99US-0137222.
PR 03-JUN-1999; 99US-0137528.
PR 04-JUN-1999; 99US-0137502.
PR 07-JUN-1999; 99US-0137724.
PR 08-JUN-1999; 99US-0138094.
PR 10-JUN-1999; 99US-0138540.
PR 14-JUN-1999; 99US-0138847.
PR 16-JUN-1999; 99US-0139119.
PR 16-JUN-1999; 99US-0139452.
PR 17-JUN-1999; 99US-0139453.
PR 18-JUN-1999; 99US-0139454.
PR 18-JUN-1999; 99US-0139455.
PR 18-JUN-1999; 99US-0139456.
PR 18-JUN-1999; 99US-0139457.
PR 18-JUN-1999; 99US-0139458.
PR 18-JUN-1999; 99US-0139459.
PR 18-JUN-1999; 99US-0139460.
PR 18-JUN-1999; 99US-0139461.
PR 18-JUN-1999; 99US-0139462.
PR 18-JUN-1999; 99US-0139463.
PR 18-JUN-1999; 99US-0139750.
PR 18-JUN-1999; 99US-0139763.
PR 21-JUN-1999; 99US-0139817.
PR 22-JUN-1999; 99US-0139889.
PR 23-JUN-1999; 99US-0140353.
PR 23-JUN-1999; 99US-0140354.
PR 24-JUN-1999; 99US-0140695.
PR 28-JUN-1999; 99US-0140823.
PR 29-JUN-1999; 99US-0140991.
PR 30-JUN-1999; 99US-0141287.
PR 01-JUL-1999; 99US-0141842.
PR 01-JUL-1999; 99US-0142154.
PR 02-JUL-1999; 99US-0142055.
PR 06-JUL-1999; 99US-0142390.
PR 08-JUL-1999; 99US-0142803.
PR 09-JUL-1999; 99US-0142920.
PR 12-JUL-1999; 99US-0142977.
PR 13-JUL-1999; 99US-0143542.
PR 14-JUL-1999; 99US-0143624.
PR 15-JUL-1999; 99US-0144005.
PR 16-JUL-1999; 99US-0144085.
PR 16-JUL-1999; 99US-0144086.
PR 19-JUL-1999; 99US-0144325.
PR 19-JUL-1999; 99US-0144331.
PR 19-JUL-1999; 99US-0144332.
PR 19-JUL-1999; 99US-0144333.
PR 19-JUL-1999; 99US-0144334.
PR 19-JUL-1999; 99US-0144335.
PR 20-JUL-1999; 99US-0144352.
PR 20-JUL-1999; 99US-0144632.
PR 20-JUL-1999; 99US-0144884.
PR 21-JUL-1999; 99US-0144814.
PR 21-JUL-1999; 99US-0145086.
PR 21-JUL-1999; 99US-0145088.
PR 21-JUL-1999; 99US-0145089.

```

PR 22-JUL-1999; 99US-0145085.
PR 22-JUL-1999; 99US-0145087.
PR 22-JUL-1999; 99US-0145089.
PR 22-JUL-1999; 99US-0145192.
PR 23-JUL-1999; 99US-0145192.
PR 23-JUL-1999; 99US-0145192.
PR 23-JUL-1999; 99US-0145218.
PR 26-JUL-1999; 99US-0145224.
PR 27-JUL-1999; 99US-0145226.
PR 27-JUL-1999; 99US-0145913.
PR 27-JUL-1999; 99US-0145918.
PR 28-JUL-1999; 99US-0145919.
PR 02-AUG-1999; 99US-0145951.
PR 02-AUG-1999; 99US-0146386.
PR 02-AUG-1999; 99US-0146388.
PR 03-AUG-1999; 99US-0146389.
PR 04-AUG-1999; 99US-0147038.
PR 04-AUG-1999; 99US-0147204.
PR 05-AUG-1999; 99US-0147302.
PR 05-AUG-1999; 99US-0147192.
PR 06-AUG-1999; 99US-0147260.
PR 06-AUG-1999; 99US-0147303.
PR 09-AUG-1999; 99US-0147416.
PR 09-AUG-1999; 99US-0147493.
PR 10-AUG-1999; 99US-0147935.
PR 11-AUG-1999; 99US-0148171.
PR 12-AUG-1999; 99US-0148319.
PR 13-AUG-1999; 99US-0148319.
PR 13-AUG-1999; 99US-0148565.
PR 16-AUG-1999; 99US-0148684.
PR 17-AUG-1999; 99US-0149368.
PR 18-AUG-1999; 99US-0149175.
PR 20-AUG-1999; 99US-0149426.
PR 20-AUG-1999; 99US-0149722.
PR 20-AUG-1999; 99US-0149723.
PR 23-AUG-1999; 99US-0149902.
PR 23-AUG-1999; 99US-0149902.
PR 25-AUG-1999; 99US-0149930.
PR 26-AUG-1999; 99US-0150566.
PR 27-AUG-1999; 99US-0150884.
PR 27-AUG-1999; 99US-0151065.
PR 27-AUG-1999; 99US-0151066.
PR 30-AUG-1999; 99US-0151080.
PR 30-AUG-1999; 99US-0151303.
PR 31-AUG-1999; 99US-0151438.
PR 01-SEP-1999; 99US-0151930.
PR 07-SEP-1999; 99US-0152363.
PR 10-SEP-1999; 99US-0153070.
PR 13-SEP-1999; 99US-0153758.
PR 16-SEP-1999; 99US-0154018.
PR 20-SEP-1999; 99US-0154039.
PR 22-SEP-1999; 99US-0154779.
PR 23-SEP-1999; 99US-0155139.
PR 24-SEP-1999; 99US-0155486.
PR 28-SEP-1999; 99US-0155659.
PR 29-SEP-1999; 99US-0156458.
PR 04-OCT-1999; 99US-0156596.
PR 05-OCT-1999; 99US-0157117.
PR 06-OCT-1999; 99US-0157753.
PR 07-OCT-1999; 99US-0157865.
PR 08-OCT-1999; 99US-0158029.
PR 12-OCT-1999; 99US-0158232.
PR 13-OCT-1999; 99US-0158369.
PR 13-OCT-1999; 99US-0159293.
PR 13-OCT-1999; 99US-0159294.
PR 13-OCT-1999; 99US-0159295.
PR 14-OCT-1999; 99US-0159329.
PR 14-OCT-1999; 99US-0159330.
PR 14-OCT-1999; 99US-0159331.
PR 14-OCT-1999; 99US-0159637.
PR 14-OCT-1999; 99US-0159638.
PR 18-OCT-1999; 99US-0159584.
PR 21-OCT-1999; 99US-0160741.
PR 21-OCT-1999; 99US-0160767.

```

```

PR 21-OCT-1999; 99US-0160768.
PR 21-OCT-1999; 99US-0160770.
PR 21-OCT-1999; 99US-0160814.
PR 21-OCT-1999; 99US-0160815.
PR 22-OCT-1999; 99US-0160815.
PR 22-OCT-1999; 99US-0160815.
PR 22-OCT-1999; 99US-0160980.
PR 22-OCT-1999; 99US-0160981.
PR 22-OCT-1999; 99US-0160989.
PR 25-OCT-1999; 99US-0161404.
PR 25-OCT-1999; 99US-0161405.
PR 25-OCT-1999; 99US-0161406.
PR 26-OCT-1999; 99US-0161359.
PR 26-OCT-1999; 99US-0161360.
PR 26-OCT-1999; 99US-0161361.
PR 28-OCT-1999; 99US-0161920.
PR 28-OCT-1999; 99US-0161992.
PR 28-OCT-1999; 99US-0161993.
PR 29-OCT-1999; 99US-0162142.

```

Query Match 27.1%; Score 726; DB 21; Length 461;
 Best Local Similarity 32.3%; Pred. No. 1.5e-55;
 Matches 166; Conservative 91; Mismatches 181; Indels 76; Gaps 10;

```

QY 12 LLLPLSLILLFLILLLKRRN--RKRRLPPECKSGWPLGTTIGYLPYATTTGDEMQ 69
DB 7 LLITVSLCSAL-----RNMQRRTKNGLPPTGMWPIFGTTFELKQGP-----NFMK 55
QY 70 QHVSXKYGIYRSNLFEGEPTIVSADAGLNFILONEGRFLFECYSPRSIGIIGKSMVLV 129
DB 56 NQRLRYGSPFKSHLGCPTLLISMDSVNNKILINKESKGLVPGYPOSMIDILGTCMAVAH 115
QY 130 GDMHRMRKSIISLNFSLHARLRTLLKDYERATLFLVDSWOONSIFSADDEAKKFTF--- 185
DB 116 GSHRLMRGSLLSLISSTMMRDHILPKVDHFMRSYLDQMNLEVIDIDKTKRHMAFSSL 175
QY 186 -----NIMAKHIMSMDPGEETEOLEKKEVTFMKGVASAPLNPETAHKALOSATYLLK 240
DB 176 TQIAGLTKRPFY-----EKKTIAPFKLVGTLSVPIDLPPTNRCGIQARNIDR 225
QY 241 FIERKMEERKLDIKEEDDEEERKTEDEAEMSKSDHVRKQRTDDILGMLKHSN---L 296
DB 226 LLRELMQERR-----DSGE-----TFDMGLCYLKKKGNRPL 258
QY 297 STEQILDLLSLFAGHETSSVAIALAFPIQACPKAVEELREEHLEIARAKKELGSEL 356
DB 259 TDEIRDOVITLLISGTYETVSTSMALRKYLDHPRKALQELRAHLEAFREKRQ--DEPL 316
QY 357 NMDDYKKMDFTQCVINETRLGNVYRFLRRALKDOVRKYGDIIPSGKVLPIVSAVLDN 416
DB 317 GLEDVKSMKFTRAVIYETSLATVNGVLRKTTDLEINGLILKGRMRYTYTREINYDA 376
QY 417 SRYDQPLNFENRMOQOONNGASSSGSESTWGNMYPFGGPGPLCAGSELAKLEMAVFI 476
DB 377 NLYEDPLLENFMRMKKSLSQ-----NSCFVFGGGRFLCGKELGIYEISFL 425
QY 477 HHLVLFKFWMLAEDDOPFAFPFVDFPNGILPIRVS 510
DB 426 HYFVTRRMEBIGDELMVFPFVAFPGFHRLIS 459

```

RESULT 12
 AAG46489 standard; Protein; 462 AA.
 AAG46489;
 18-OCT-2000 (first entry)
 Arabidopsis thaliana protein fragment SEQ ID NO: 58493.
 Protein identification: signal transduction pathway; metabolic pathway;
 hybridisation assay; genetic mapping; gene expression control; promoter;
 termination sequence.
 XX

OS Arabidopsis thaliana.
XX EP1033405-A2.
XX
PD 06-SEP-2000.
XX
PF 25-FEB-2000; 2000EP-0301439.
XX
PR 25-FEB-1999; 99US-0121825.
PR 05-MAR-1999; 99US-0123180.
PR 09-MAR-1999; 99US-0123548.
PR 23-MAR-1999; 99US-0125788.
PR 25-MAR-1999; 99US-0126264.
PR 29-MAR-1999; 99US-0126785.
PR 01-APR-1999; 99US-0127462.
PR 06-APR-1999; 99US-0128234.
PR 08-APR-1999; 99US-0128714.
PR 16-APR-1999; 99US-0129845.
PR 19-APR-1999; 99US-0130077.
PR 21-APR-1999; 99US-0130449.
PR 23-APR-1999; 99US-0130510.
PR 28-APR-1999; 99US-0130891.
PR 30-APR-1999; 99US-0131449.
PR 30-APR-1999; 99US-0132048.
PR 04-MAY-1999; 99US-0132407.
PR 05-MAY-1999; 99US-0132485.
PR 06-MAY-1999; 99US-0132486.
PR 07-MAY-1999; 99US-0132487.
PR 11-MAY-1999; 99US-0132863.
PR 14-MAY-1999; 99US-0134256.
PR 14-MAY-1999; 99US-0134218.
PR 14-MAY-1999; 99US-0134219.
PR 14-MAY-1999; 99US-0134221.
PR 14-MAY-1999; 99US-0134370.
PR 18-MAY-1999; 99US-0134768.
PR 19-MAY-1999; 99US-0134941.
PR 20-MAY-1999; 99US-0135124.
PR 21-MAY-1999; 99US-0135353.
PR 25-MAY-1999; 99US-0135629.
PR 27-MAY-1999; 99US-0136021.
PR 28-MAY-1999; 99US-0136392.
PR 01-JUN-1999; 99US-0136782.
PR 03-JUN-1999; 99US-0137222.
PR 04-JUN-1999; 99US-0137502.
PR 07-JUN-1999; 99US-0137724.
PR 08-JUN-1999; 99US-0138094.
PR 10-JUN-1999; 99US-0138540.
PR 14-JUN-1999; 99US-0138847.
PR 16-JUN-1999; 99US-0139119.
PR 16-JUN-1999; 99US-0139452.
PR 17-JUN-1999; 99US-0139453.
PR 18-JUN-1999; 99US-0139454.
PR 18-JUN-1999; 99US-0139455.
PR 18-JUN-1999; 99US-0139456.
PR 18-JUN-1999; 99US-0139457.
PR 18-JUN-1999; 99US-0139458.
PR 18-JUN-1999; 99US-0139459.
PR 18-JUN-1999; 99US-0139460.
PR 18-JUN-1999; 99US-0139461.
PR 18-JUN-1999; 99US-0139462.
PR 18-JUN-1999; 99US-0139463.
PR 18-JUN-1999; 99US-0139750.
PR 18-JUN-1999; 99US-0139763.
PR 21-JUN-1999; 99US-0139817.
PR 22-JUN-1999; 99US-0139839.
PR 23-JUN-1999; 99US-0140353.
PR 23-JUN-1999; 99US-0140354.
PR 24-JUN-1999; 99US-0140695.
PR 28-JUN-1999; 99US-0140823.
PR 29-JUN-1999; 99US-0140991.

PR 30-JUN-1999; 99US-0141287.
PR 01-JUL-1999; 99US-0141842.
PR 01-JUL-1999; 99US-0142154.
PR 02-JUL-1999; 99US-0142055.
PR 06-JUL-1999; 99US-0142390.
PR 08-JUL-1999; 99US-0142803.
PR 09-JUL-1999; 99US-0142920.
PR 12-JUL-1999; 99US-0142977.
PR 13-JUL-1999; 99US-0143544.
PR 14-JUL-1999; 99US-0143624.
PR 15-JUL-1999; 99US-0144005.
PR 16-JUL-1999; 99US-0144085.
PR 16-JUL-1999; 99US-0144086.
PR 19-JUL-1999; 99US-0144325.
PR 19-JUL-1999; 99US-0144331.
PR 19-JUL-1999; 99US-0144332.
PR 19-JUL-1999; 99US-0144333.
PR 19-JUL-1999; 99US-0144334.
PR 19-JUL-1999; 99US-0144335.
PR 20-JUL-1999; 99US-0144632.
PR 20-JUL-1999; 99US-0144637.
PR 21-JUL-1999; 99US-0144814.
PR 21-JUL-1999; 99US-0145086.
PR 21-JUL-1999; 99US-0145088.
PR 22-JUL-1999; 99US-0145087.
PR 22-JUL-1999; 99US-0145089.
PR 22-JUL-1999; 99US-0145192.
PR 23-JUL-1999; 99US-0145145.
PR 23-JUL-1999; 99US-0145218.
PR 23-JUL-1999; 99US-0145224.
PR 26-JUL-1999; 99US-0145276.
PR 27-JUL-1999; 99US-0145213.
PR 27-JUL-1999; 99US-0145918.
PR 27-JUL-1999; 99US-0145919.
PR 28-JUL-1999; 99US-0145951.
PR 02-AUG-1999; 99US-0146386.
PR 02-AUG-1999; 99US-0146388.
PR 03-AUG-1999; 99US-0146389.
PR 04-AUG-1999; 99US-0147038.
PR 04-AUG-1999; 99US-0147204.
PR 04-AUG-1999; 99US-0147302.
PR 05-AUG-1999; 99US-0147192.
PR 05-AUG-1999; 99US-0147260.
PR 06-AUG-1999; 99US-0147303.
PR 06-AUG-1999; 99US-0147416.
PR 09-AUG-1999; 99US-0147493.
PR 09-AUG-1999; 99US-0147933.
PR 10-AUG-1999; 99US-0148171.
PR 11-AUG-1999; 99US-0148319.
PR 12-AUG-1999; 99US-0148341.
PR 13-AUG-1999; 99US-0148565.
PR 13-AUG-1999; 99US-0148684.
PR 16-AUG-1999; 99US-0149368.
PR 17-AUG-1999; 99US-0149175.
PR 18-AUG-1999; 99US-0149426.
PR 20-AUG-1999; 99US-0149422.
PR 20-AUG-1999; 99US-0149723.
PR 20-AUG-1999; 99US-0149929.
PR 23-AUG-1999; 99US-0149902.
PR 23-AUG-1999; 99US-0149930.
PR 25-AUG-1999; 99US-0150566.
PR 26-AUG-1999; 99US-0150884.
PR 27-AUG-1999; 99US-0151065.
PR 27-AUG-1999; 99US-0151066.
PR 27-AUG-1999; 99US-0151080.
PR 30-AUG-1999; 99US-0151303.
PR 31-AUG-1999; 99US-0151438.
PR 01-SEP-1999; 99US-0151930.
PR 07-SEP-1999; 99US-0152363.
PR 10-SEP-1999; 99US-0153070.
PR 13-SEP-1999; 99US-0153758.

Query Match	27.1%	Score 726;	DB 21;	Length 462;
Best Local Similarity	32.3%	Pred. No. 1.5e-55;		
Matches 166;	Conservative 91;	Mismatches 181;	Indels 76;	Gaps 101;
QY	12	LLLPFLSLFLFLILKRRN--KRTRNLPGRSGMFLGEGTGLKPRYATILGLDMQ	69	
Db	8	LLIVLISCALL-----RNMQRYYTKKGLPPRMGMWIFSEETEEFLQGP----	56	NEMR
QY	70	QHWKYGKIKYRSINLFGEPPIVSADAGNLFILIONEGRLFECSYRSIGILGKMSLVV	129	
Db	57	NQRLRYGSPFKSHLGGPTLISMDSQEVNRYILKKNSKGLVGYGQSMILDIIGTCMAVH	116	
QY	130	GDHNRDMRSISLNLSHARLFTILIKDVERHTFLVDSWOONSIFSADEAKKETP----	185	
Db	117	GSSHRIMRGSLISLISSTMMDHILIPYDHFMRSYLDQWNELEVLDIDQTKRHNAFLSSL	176	
QY	186	-----NIMAKHMSMDPGEETEDLKEEYVTFMGVYSAPLNDEGTAYAKKLOSFATILK	240	
Db	177	TQIAGNLRKRPV-----EEFKAFELVYGLTSLVPIDLPGTNRCGIOARNNIDR	226	
QY	241	FIERKMEERLDIKEEDQEEEEVTEDEAEKSKSDHYRKQRTDDDLIGWVLHNSN----	296	L
Db	227	LLELWQERR-----DSGE-----TFYDMIGILKKKBSGRNYP	259	
QY	297	STEOIIDLILSLFAGHETSSVALAIAIFLQACPRAVEELREHLEIARAKKETGESEL	356	
Db	260	TDEIRIQOYVITIIYSGEYVSTSMALKYLIHDHPKQLOELRAEHLAIFREKRKQ--DEPL	317	
QY	357	NMDYKMDTQCVINTEILGNVYRFLHKKALKDVRYKGYDIPDSGWKVFVISAVALHDN	416	

Dd		318 GLEDVSKSMKFTRAVIYETSRATIIVGVLKTRDLEINQYLIPKGRIITYPREINYDA 377
Oy	417	KRYDPNLFNPWRMOOQNNGASSSGSESTGCNNTPYEGGGPRLCAGSELANKEMAVFI 476
Dd	378	NLYEDELFINPWPMWMMKSLSLSEQ-----NSCFVEFGGTRLCPGKELGIVEISSFL 426
Oy	477	HHVLKFENNELAEDDDQFAFPFVPDFPNGLPIRYS 510
Dd	427	HVFTRYRMWEELIGDELMVFPFYFAFKGFHLRIS 460
 RESULT 13 AAG46491 ID AAG46491 standard; Protein; 465 AA.		
XX	AAG46491;	
AC		
XX		
DT	18-OCT-2000 (first entry)	
XX		
DE	Arabidopsis thaliana protein fragment SEQ ID NO: 58495.	
XX		
XX	Protein identification; signal transduction pathway; metabolic pathway; KW hybridisation assay; genetic mapping; gene expression control; promoter; KM termination sequence.	
OS	Arabidopsis thaliana.	
XX		
PN	EPI033405-A2.	
PD		
XX	06-SEP-2000.	
PF	25-FEB-2000; 200EP-0301439.	
XX		
PR	25-FEB-1999; 99US-0121825.	
PR	05-MAR-1999; 99US-0122180.	
PR	09-MAR-1999; 99US-0123548.	
PR	23-MAR-1999; 99US-0125788.	
PR	25-MAR-1999; 99US-0126264.	
PR	29-MAR-1999; 99US-0126785.	
PR	01-APR-1999; 99US-0127462.	
PR	06-APR-1999; 99US-0128234.	
PR	08-APR-1999; 99US-0128714.	
PR	16-APR-1999; 99US-0128845.	
PR	19-APR-1999; 99US-0130077.	
PR	21-APR-1999; 99US-0130449.	
PR	23-APR-1999; 99US-0130510.	
PR	23-APR-1999; 99US-0130891.	
PR	28-APR-1999; 99US-0131449.	
PR	30-APR-1999; 99US-0132048.	
PR	30-APR-1999; 99US-0132407.	
PR	04-MAY-1999; 99US-0132484.	
PR	05-MAY-1999; 99US-0132485.	
PR	06-MAY-1999; 99US-0132486.	
PR	06-MAY-1999; 99US-0132487.	
PR	07-MAY-1999; 99US-0132863.	
PR	11-MAY-1999; 99US-0134256.	
PR	14-MAY-1999; 99US-0134219.	
PR	14-MAY-1999; 99US-0134221.	
PR	14-MAY-1999; 99US-0134370.	
PR	18-MAY-1999; 99US-0134768.	
PR	19-MAY-1999; 99US-0134941.	
PR	20-MAY-1999; 99US-0135124.	
PR	21-MAY-1999; 99US-0135353.	
PR	24-MAY-1999; 99US-0135629.	
PR	25-MAY-1999; 99US-0136021.	
PR	27-MAY-1999; 99US-0136392.	
PR	28-MAY-1999; 99US-0136782.	
PR	01-JUN-1999; 99US-0137222.	
PR	03-JUN-1999; 99US-0137528.	
PR	04-JUN-1999; 99US-0137502.	
PR	07-JUN-1999; 99US-0137724.	
PR	08-JUN-1999; 99US-0138094.	

PR 10-JUN-1999; 99US-0138540.
PR 10-JUN-1999; 99US-0138847.
PR 14-JUN-1999; 99US-0139119.
PR 16-JUN-1999; 99US-0139452.
PR 16-JUN-1999; 99US-0139453.
PR 17-JUN-1999; 99US-0139492.
PR 18-JUN-1999; 99US-0139454.
PR 18-JUN-1999; 99US-0139455.
PR 18-JUN-1999; 99US-0139456.
PR 18-JUN-1999; 99US-0139457.
PR 18-JUN-1999; 99US-0139458.
PR 18-JUN-1999; 99US-0139459.
PR 18-JUN-1999; 99US-0139460.
PR 18-JUN-1999; 99US-0139461.
PR 18-JUN-1999; 99US-0139462.
PR 18-JUN-1999; 99US-0139463.
PR 18-JUN-1999; 99US-0139750.
PR 18-JUN-1999; 99US-0139763.
PR 21-JUN-1999; 99US-0139817.
PR 22-JUN-1999; 99US-0139839.
PR 23-JUN-1999; 99US-0140353.
PR 23-JUN-1999; 99US-0140354.
PR 24-JUN-1999; 99US-0140695.
PR 28-JUN-1999; 99US-0140823.
PR 29-JUN-1999; 99US-0140981.
PR 30-JUN-1999; 99US-0141287.
PR 01-JUL-1999; 99US-0141842.
PR 02-JUL-1999; 99US-0142154.
PR 02-JUL-1999; 99US-0142055.
PR 06-JUL-1999; 99US-0142390.
PR 08-JUL-1999; 99US-0142803.
PR 09-JUL-1999; 99US-0142920.
PR 12-JUL-1999; 99US-0142977.
PR 13-JUL-1999; 99US-0143542.
PR 14-JUL-1999; 99US-0143624.
PR 15-JUL-1999; 99US-0144005.
PR 16-JUL-1999; 99US-0144085.
PR 16-JUL-1999; 99US-0144086.
PR 19-JUL-1999; 99US-0144325.
PR 19-JUL-1999; 99US-0144331.
PR 19-JUL-1999; 99US-0144332.
PR 19-JUL-1999; 99US-0144333.
PR 19-JUL-1999; 99US-0144334.
PR 19-JUL-1999; 99US-0144335.
PR 20-JUL-1999; 99US-0144352.
PR 20-JUL-1999; 99US-0144632.
PR 20-JUL-1999; 99US-0144884.
PR 21-JUL-1999; 99US-0144814.
PR 21-JUL-1999; 99US-0145086.
PR 21-JUL-1999; 99US-0145088.
PR 22-JUL-1999; 99US-0145085.
PR 22-JUL-1999; 99US-0145087.
PR 22-JUL-1999; 99US-0145089.
PR 22-JUL-1999; 99US-0145192.
PR 23-JUL-1999; 99US-0145145.
PR 23-JUL-1999; 99US-0145218.
PR 23-JUL-1999; 99US-0145224.
PR 26-JUL-1999; 99US-0145276.
PR 27-JUL-1999; 99US-0145913.
PR 27-JUL-1999; 99US-0145918.
PR 27-JUL-1999; 99US-0145919.
PR 28-JUL-1999; 99US-0145951.
PR 02-AUG-1999; 99US-0146386.
PR 02-AUG-1999; 99US-0146388.
PR 03-AUG-1999; 99US-0146389.
PR 03-AUG-1999; 99US-0147038.
PR 04-AUG-1999; 99US-0147204.
PR 04-AUG-1999; 99US-0147302.
PR 05-AUG-1999; 99US-0147192.
PR 05-AUG-1999; 99US-0147260.
PR 06-AUG-1999; 99US-0147303.
PR 06-AUG-1999; 99US-0147416.
PR 09-AUG-1999; 99US-0147493.

PR 09-AUG-1999; 99US-0147935.
PR 10-AUG-1999; 99US-0148171.
PR 11-AUG-1999; 99US-0148319.
PR 12-AUG-1999; 99US-0148341.
PR 13-AUG-1999; 99US-0148565.
PR 13-AUG-1999; 99US-0148684.
PR 16-AUG-1999; 99US-0149368.
PR 17-AUG-1999; 99US-0149175.
PR 18-AUG-1999; 99US-0149175.
PR 20-AUG-1999; 99US-0149722.
PR 20-AUG-1999; 99US-0149723.
PR 20-AUG-1999; 99US-0149928.
PR 23-AUG-1999; 99US-0149902.
PR 23-AUG-1999; 99US-0149930.
PR 25-AUG-1999; 99US-0150566.
PR 26-AUG-1999; 99US-0150884.
PR 27-AUG-1999; 99US-0151065.
PR 27-AUG-1999; 99US-0151066.
PR 27-AUG-1999; 99US-0151080.
PR 30-AUG-1999; 99US-0151303.
PR 31-AUG-1999; 99US-0151438.
PR 01-SEP-1999; 99US-0151930.
PR 07-SEP-1999; 99US-0152363.
PR 10-SEP-1999; 99US-0153070.
PR 13-SEP-1999; 99US-0153758.
PR 15-SEP-1999; 99US-0154018.
PR 16-SEP-1999; 99US-0154039.
PR 20-SEP-1999; 99US-0154779.
PR 22-SEP-1999; 99US-0155139.
PR 23-SEP-1999; 99US-0155486.
PR 24-SEP-1999; 99US-0155659.
PR 28-SEP-1999; 99US-0156458.
PR 29-SEP-1999; 99US-0156596.
PR 04-OCT-1999; 99US-0157117.
PR 05-OCT-1999; 99US-0157753.
PR 06-OCT-1999; 99US-0157865.
PR 07-OCT-1999; 99US-0158028.
PR 08-OCT-1999; 99US-0158232.
PR 12-OCT-1999; 99US-0158369.
PR 13-OCT-1999; 99US-0159293.
PR 13-OCT-1999; 99US-0159294.
PR 13-OCT-1999; 99US-0159295.
PR 14-OCT-1999; 99US-0159329.
PR 14-OCT-1999; 99US-0159330.
PR 14-OCT-1999; 99US-0159331.
PR 14-OCT-1999; 99US-0159637.
PR 14-OCT-1999; 99US-0159638.
PR 18-OCT-1999; 99US-0159584.
PR 21-OCT-1999; 99US-0160741.
PR 21-OCT-1999; 99US-0160767.
PR 21-OCT-1999; 99US-0160768.
PR 21-OCT-1999; 99US-0160770.
PR 21-OCT-1999; 99US-0160814.
PR 21-OCT-1999; 99US-0160815.
PR 22-OCT-1999; 99US-0160960.
PR 22-OCT-1999; 99US-0160981.
PR 22-OCT-1999; 99US-0160989.
PR 25-OCT-1999; 99US-0161404.
PR 25-OCT-1999; 99US-0161405.
PR 25-OCT-1999; 99US-0161406.
PR 26-OCT-1999; 99US-0161359.
PR 26-OCT-1999; 99US-0161360.
PR 26-OCT-1999; 99US-0161361.
PR 28-OCT-1999; 99US-0161920.
PR 28-OCT-1999; 99US-0161992.
PR 28-OCT-1999; 99US-0161993.
PR 29-OCT-1999; 99US-0162142.

Query Match 27.1%; Score 726; DB 21; Length 465;
Best Local Similarity 32.3%; Pred. No. 1.5e-55;
Matches 166; Conservative 91; Mismatches 181; Indels 76; Gaps 10;
QY 12 LLLPILLSLLFLILKRRN--RKTRENLPPGKSGWPFGLGTYLGYLKPATTLGDFMQ 69

Db 11 ILIIVSLCSALL-----RWNOHRYTKNGLPPTGMGPIFGETTELKQGP-----NFWR 59
QY 70 QHVSXKGIYRNLGEPITVASDAGLNFTILQNEBRLEECYSPRSIGILKMSMLV 129
Db 60 NORLKGSEFKSHLIGCPPLISMDEVNXYILKNEKSGLVPGYPSMDLITCQNAVAH 119
QY 130 GDMHRDMSISLNFLSHARLFTLLKDYERHTLFVLDSDQONSIFSADQAKKFFP-----185
Db 120 GSSHLMGSLSLSLSTMMRHILPKVDHFKRSYLDONNELEVIDIQKTHMAFLSL 179
QY 186 -----NLAKHITMSDPEEETEQLKKEYVTPEMGVSNPLNPGTAYHKAQSRATILK 240
Db 180 TQIAGNLKRPV-----EEFKTAFKLVYGTLSVIDLPNTNRCGIQARNMIDR 229
QY 241 FIERKEERKLDIKKEDEEVEKTEDEAKMSKSHVRKQRDDDLGVLAKHSN----L 296
Db 230 LIRELMQERR-----DSGE-----FTDMLGYLMKKEGNRPYL 262
QY 297 STEQILDLILSLFAGHETSVAIALAIFLQACPKAVBELREHELEIRAKKELGESL 356
Db 263 TDEEIRDOVVTILYSGYEVSTSMALKYLHDHRAQLOELAEHLAFERRKQ--DEPL 320
QY 357 NMDDVKKDPFQCVINETIRLCNVYRFLHRAKADVRKGYDIPSGWVLPVISAVIDLN 416
Db 321 GLEDVKSMEFTRAVYETSRLATIYNGVLKRTDLEINGYLIPKGMRIYVYTRREINYA 380
QY 417 SRYDQNLFPNPMOONNASSSGSFSTGNNYMPGSGPRICAGSELAKLEMAVPI 476
Db 381 NIEYDPLIFNPRMKKSLESO-----NSCFYFGSGTRICPGKELIIVETISPL 429
QY 477 HHVLKFNWELAEDDQPEAFPPVDPNGLPIRVNS 510
Db 430 HFFVTRWEEIGDELMVFPVFAKPGFHLRIS 463

RESULT 14
AAG11836
ID AAG11836 standard; Protein; 461 AA.
XX AAG11836;
AC
XX
XX
DT 17-OCT-2000 (first entry)
XX
DE Arabidopsis thaliana protein fragment SEQ ID NO: 10708.
XX
KW Protein identification; signal transduction pathway; metabolic pathway;
KW hybridisation assay; genetic mapping; gene expression control; promoter;
KW termination sequence.
XX
OS Arabidopsis thaliana.
XX
XX
PN EP1033405-A2.
XX
PD 06-SEP-2000.
XX
PF 25-FEB-2000; 2000EP-0301439.
XX
PR 25-FEB-1999; 990S-0121825.
PR 05-MAR-1999; 990S-0123180.
PR 09-MAR-1999; 990S-0123548.
PR 23-MAR-1999; 990S-0123788.
PR 25-MAR-1999; 990S-0126264.
PR 29-MAR-1999; 990S-0126785.
PR 01-APR-1999; 990S-0127462.
PR 06-APR-1999; 990S-0128234.
PR 08-APR-1999; 990S-0128714.
PR 16-APR-1999; 990S-0129845.
PR 19-APR-1999; 990S-0130077.
PR 21-APR-1999; 990S-0130449.
PR 23-APR-1999; 990S-0130510.
PR 23-APR-1999; 990S-0130891.
PR 28-APR-1999; 990S-0131445.

PR 30-APR-1999; 990S-0132048.
PR 30-APR-1999; 990S-0132407.
PR 04-MAY-1999; 990S-0132484.
PR 05-MAY-1999; 990S-0132485.
PR 06-MAY-1999; 990S-0132486.
PR 06-MAY-1999; 990S-0132487.
PR 07-MAY-1999; 990S-0132488.
PR 11-MAY-1999; 990S-0134256.
PR 14-MAY-1999; 990S-0134218.
PR 14-MAY-1999; 990S-0134219.
PR 14-MAY-1999; 990S-0134221.
PR 14-MAY-1999; 990S-0134370.
PR 18-MAY-1999; 990S-0134768.
PR 19-MAY-1999; 990S-0134941.
PR 20-MAY-1999; 990S-0135124.
PR 21-MAY-1999; 990S-0135124.
PR 21-MAY-1999; 990S-0135124.
PR 24-MAY-1999; 990S-0135629.
PR 25-MAY-1999; 990S-0136021.
PR 27-MAY-1999; 990S-0136392.
PR 28-MAY-1999; 990S-0136782.
PR 01-JUN-1999; 990S-0137222.
PR 03-JUN-1999; 990S-0137528.
PR 04-JUN-1999; 990S-0137502.
PR 07-JUN-1999; 990S-0137724.
PR 08-JUN-1999; 990S-0138094.
PR 10-JUN-1999; 990S-0138540.
PR 10-JUN-1999; 990S-0138847.
PR 14-JUN-1999; 990S-0139119.
PR 16-JUN-1999; 990S-0139452.
PR 16-JUN-1999; 990S-0139453.
PR 17-JUN-1999; 990S-0139454.
PR 18-JUN-1999; 990S-0139492.
PR 18-JUN-1999; 990S-0139454.
PR 18-JUN-1999; 990S-0139455.
PR 18-JUN-1999; 990S-0139456.
PR 18-JUN-1999; 990S-0139457.
PR 18-JUN-1999; 990S-0139458.
PR 18-JUN-1999; 990S-0139459.
PR 18-JUN-1999; 990S-0139460.
PR 18-JUN-1999; 990S-0139461.
PR 18-JUN-1999; 990S-0139462.
PR 18-JUN-1999; 990S-0139463.
PR 18-JUN-1999; 990S-0139463.
PR 18-JUN-1999; 990S-0139750.
PR 21-JUN-1999; 990S-0139763.
PR 21-JUN-1999; 990S-0139817.
PR 22-JUN-1999; 990S-0139899.
PR 23-JUN-1999; 990S-0140353.
PR 23-JUN-1999; 990S-0140354.
PR 24-JUN-1999; 990S-0140695.
PR 28-JUN-1999; 990S-0140823.
PR 29-JUN-1999; 990S-0140991.
PR 30-JUN-1999; 990S-0141287.
PR 01-JUL-1999; 990S-0141842.
PR 01-JUL-1999; 990S-0142154.
PR 02-JUL-1999; 990S-0142055.
PR 06-JUL-1999; 990S-0142390.
PR 08-JUL-1999; 990S-0142803.
PR 09-JUL-1999; 990S-0142920.
PR 12-JUL-1999; 990S-0142977.
PR 13-JUL-1999; 990S-0143542.
PR 14-JUL-1999; 990S-0143624.
PR 15-JUL-1999; 990S-0144005.
PR 16-JUL-1999; 990S-0144085.
PR 16-JUL-1999; 990S-0144086.
PR 19-JUL-1999; 990S-0144325.
PR 19-JUL-1999; 990S-0144331.
PR 19-JUL-1999; 990S-0144332.
PR 19-JUL-1999; 990S-0144333.
PR 19-JUL-1999; 990S-0144334.
PR 19-JUL-1999; 990S-0144335.
PR 20-JUL-1999; 990S-0144352.
PR 20-JUL-1999; 990S-0144632.
PR 20-JUL-1999; 990S-0144884.
PR 21-JUL-1999; 990S-0144814.

```

PR 21-JUL-1999; 99US-0145086.
PR 21-JUL-1999; 99US-0145088.
PR 22-JUL-1999; 99US-0145085.
PR 22-JUL-1999; 99US-0145087.
PR 22-JUL-1999; 99US-0145089.
PR 22-JUL-1999; 99US-0145192.
PR 23-JUL-1999; 99US-0145145.
PR 23-JUL-1999; 99US-0145218.
PR 23-JUL-1999; 99US-0145224.
PR 26-JUL-1999; 99US-0145276.
PR 27-JUL-1999; 99US-0145913.
PR 27-JUL-1999; 99US-0145918.
PR 27-JUL-1999; 99US-0145919.
PR 28-JUL-1999; 99US-0145951.
PR 02-AUG-1999; 99US-0146386.
PR 02-AUG-1999; 99US-0146388.
PR 02-AUG-1999; 99US-0146389.
PR 03-AUG-1999; 99US-0147038.
PR 04-AUG-1999; 99US-0147204.
PR 04-AUG-1999; 99US-0147302.
PR 05-AUG-1999; 99US-0147192.
PR 05-AUG-1999; 99US-0147260.
PR 06-AUG-1999; 99US-0147303.
PR 06-AUG-1999; 99US-0147416.
PR 09-AUG-1999; 99US-0147493.
PR 09-AUG-1999; 99US-0147935.
PR 10-AUG-1999; 99US-0148171.
PR 11-AUG-1999; 99US-0148319.
PR 12-AUG-1999; 99US-0148341.
PR 13-AUG-1999; 99US-0148565.
PR 13-AUG-1999; 99US-0148684.
PR 16-AUG-1999; 99US-0149368.
PR 17-AUG-1999; 99US-0149175.
PR 18-AUG-1999; 99US-0149426.
PR 20-AUG-1999; 99US-0149722.
PR 20-AUG-1999; 99US-0149723.
PR 20-AUG-1999; 99US-0149929.
PR 23-AUG-1999; 99US-0149902.
PR 23-AUG-1999; 99US-0149930.
PR 25-AUG-1999; 99US-0150566.
PR 26-AUG-1999; 99US-0150884.
PR 27-AUG-1999; 99US-0151065.
PR 27-AUG-1999; 99US-0151066.
PR 27-AUG-1999; 99US-0151080.
PR 27-AUG-1999; 99US-0151303.
PR 30-AUG-1999; 99US-0151438.
PR 31-AUG-1999; 99US-0151930.
PR 01-SEP-1999; 99US-0151930.
PR 07-SEP-1999; 99US-0152363.
PR 10-SEP-1999; 99US-0153070.
PR 13-SEP-1999; 99US-0153758.
PR 15-SEP-1999; 99US-0154018.
PR 16-SEP-1999; 99US-0154039.
PR 20-SEP-1999; 99US-0154779.
PR 22-SEP-1999; 99US-0155139.
PR 23-SEP-1999; 99US-0155486.
PR 24-SEP-1999; 99US-0155659.
PR 28-SEP-1999; 99US-0156458.
PR 29-SEP-1999; 99US-0156596.
PR 04-OCT-1999; 99US-0157117.
PR 05-OCT-1999; 99US-0157753.
PR 06-OCT-1999; 99US-0157865.
PR 07-OCT-1999; 99US-0158029.
PR 08-OCT-1999; 99US-0158232.
PR 12-OCT-1999; 99US-0158369.
PR 13-OCT-1999; 99US-0159293.
PR 13-OCT-1999; 99US-0159294.
PR 13-OCT-1999; 99US-0159295.
PR 14-OCT-1999; 99US-0159329.
PR 14-OCT-1999; 99US-0159330.
PR 14-OCT-1999; 99US-0159331.
PR 14-OCT-1999; 99US-0159637.
PR 14-OCT-1999; 99US-0159638.
PR 18-OCT-1999; 99US-0159584.

```

```

PR 21-OCT-1999; 99US-0160741.
PR 21-OCT-1999; 99US-0160767.
PR 21-OCT-1999; 99US-0160768.
PR 21-OCT-1999; 99US-0160770.
PR 21-OCT-1999; 99US-0160814.
PR 21-OCT-1999; 99US-0160815.
PR 22-OCT-1999; 99US-0160980.
PR 22-OCT-1999; 99US-0160981.
PR 22-OCT-1999; 99US-0160989.
PR 25-OCT-1999; 99US-0161404.
PR 25-OCT-1999; 99US-0161405.
PR 25-OCT-1999; 99US-0161406.
PR 26-OCT-1999; 99US-0161359.
PR 26-OCT-1999; 99US-0161360.
PR 26-OCT-1999; 99US-0161361.
PR 28-OCT-1999; 99US-0161920.
PR 28-OCT-1999; 99US-0161922.
PR 28-OCT-1999; 99US-0161993.
PR 29-OCT-1999; 99US-0162142.

```

Query Match 26.6%; Score 712; DB 21; Length 461;
 Best Local Similarity 32.5%; Pred. No. 2,5e-54;
 Matches 164; Conservative 93; Mismatches 196; Indels 52; Gaps 8;

```

QY 12 LLLPSSLILFLILKRRN--KRTRENLPKSGMPFLGTTGYLKPYATTLGDMQ 69
DB 1 MMILGLVIVICLTALRWNNOMRYSKGLPPTGWPFIQETFEFLKGP-----DEMK 55
QY 70 QHVSRYKYKIVSNLGEPTIYSDAGLNRLFLIONGRLEFESYPSIGIGLKMGLVIV 129
DB 56 NORLRYGSFFSHLIGCPTIVSMDLERNYILMNSKGLVAGYPOSMIDLITCTINAVH 115
QY 130 GDMHRDMKISLINFSLHARLRTILKLDVERHTLFVLDSMOONSIFSADDEAKKFTFNMA 189
DB 116 GPSRLRMGSLLSISPMMKDHLPKIDDFRNVLGMDLLETVDIEKTKHMAF-LSS 174
QY 190 KHMSMDGEETEDOLKKEIYTFPKGVVSAPLNPGLTYHKALOSRAATILKTEKMEER 249
DB 175 LLOIAETLLKKEVEERYRTEFFKLVYGLTSPIDIPGTNYRSGVQARNIDRLTFLMOER 234
QY 250 KLDIKKEQDEEEVYKTEDEAMSKSDHVRKQRTDDLLGVLKHSN---LSTEQILDLI 305
DB 235 K-----ESGEFTT-----DMIGYLKKKEDNNRYLLDKETIRQV 267
QY 306 LSLFAGHETSSVAIALAIFLQACPRAVEELREHLEIARAKKELGSELMDDYKMD 365
DB 268 VTILYSGYEVSTSMALKYLDHPKALEELRHLAIRREKR--DEPLTDDIKSMK 325
QY 366 FTQCYINETLRLGNVVRFLHAKALKADVRYKGYDIPSGKVLPIVISAHLDSRYDQPNLF 425
DB 326 FTRAVIFETSRATVNGVLKRTYHDLNLNGYLIPKGRIRIYYTREIYVDSLYEDPMIF 385
QY 426 NPMRMQOONNGASSSGSFSFTWCNNYMPFGCGPPLCAGSELAKLEMAVFIHHLVKRNM 485
DB 386 NPMRMKESLESKS-----YFLDFGGVRLCCKGKELGISVSLHYFTYKRW 434
QY 486 ELAEDDOPFAFPEVDFPNGLPRIYS 510
DB 435 EENGEDKIMVFPRIVSAPKGYHLKCS 459

```

RESULT 15
 AAG11835
 ID AAG11835 standard; Protein: 462 AA.
 AC AAG11835;
 DT 17-OCT-2000 (first entry)
 XX Arabidopsis thaliana protein fragment SEQ ID NO: 10707.
 DE Protein identification: signal transduction pathway; metabolic pathway;
 KW hydriatation assay; genetic mapping; gene expression control; promoter;

termination sequence.
KM
XX
OS Arabidopsis thaliana.
PM EP1033405-A2.
XX
PD 06-SEP-2000.
XX
PF 25-FEB-2000; 2000EP-0301439.
XX
PR 25-FEB-1999; 99US-0121825.
PR 05-MAR-1999; 99US-0123180.
PR 09-MAR-1999; 99US-0123548.
PR 23-MAR-1999; 99US-0125788.
PR 25-MAR-1999; 99US-0126264.
PR 29-MAR-1999; 99US-0127485.
PR 01-APR-1999; 99US-0127462.
PR 06-APR-1999; 99US-0128234.
PR 08-APR-1999; 99US-0128714.
PR 16-APR-1999; 99US-0129845.
PR 19-APR-1999; 99US-0130077.
PR 21-APR-1999; 99US-0130449.
PR 23-APR-1999; 99US-0130510.
PR 28-APR-1999; 99US-0130891.
PR 30-APR-1999; 99US-0131449.
PR 30-APR-1999; 99US-0132048.
PR 04-MAY-1999; 99US-0132484.
PR 05-MAY-1999; 99US-0132485.
PR 06-MAY-1999; 99US-0132486.
PR 07-MAY-1999; 99US-0132487.
PR 11-MAY-1999; 99US-0132863.
PR 14-MAY-1999; 99US-0134256.
PR 14-MAY-1999; 99US-0134218.
PR 14-MAY-1999; 99US-0134219.
PR 14-MAY-1999; 99US-0134221.
PR 18-MAY-1999; 99US-0134370.
PR 19-MAY-1999; 99US-0134368.
PR 20-MAY-1999; 99US-0134941.
PR 21-MAY-1999; 99US-0135124.
PR 24-MAY-1999; 99US-0135353.
PR 25-MAY-1999; 99US-0135629.
PR 27-MAY-1999; 99US-0136021.
PR 28-MAY-1999; 99US-0136392.
PR 01-JUN-1999; 99US-0136782.
PR 03-JUN-1999; 99US-0137222.
PR 04-JUN-1999; 99US-0137502.
PR 07-JUN-1999; 99US-0137724.
PR 08-JUN-1999; 99US-0138094.
PR 10-JUN-1999; 99US-0138540.
PR 14-JUN-1999; 99US-0138847.
PR 16-JUN-1999; 99US-0139119.
PR 16-JUN-1999; 99US-0139452.
PR 17-JUN-1999; 99US-0139453.
PR 17-JUN-1999; 99US-0139492.
PR 18-JUN-1999; 99US-0139454.
PR 18-JUN-1999; 99US-0139455.
PR 18-JUN-1999; 99US-0139456.
PR 18-JUN-1999; 99US-0139457.
PR 18-JUN-1999; 99US-0139458.
PR 18-JUN-1999; 99US-0139459.
PR 18-JUN-1999; 99US-0139460.
PR 18-JUN-1999; 99US-0139461.
PR 18-JUN-1999; 99US-0139462.
PR 18-JUN-1999; 99US-0139463.
PR 18-JUN-1999; 99US-0139750.
PR 18-JUN-1999; 99US-0139763.
PR 21-JUN-1999; 99US-0139817.
PR 22-JUN-1999; 99US-0139899.
PR 23-JUN-1999; 99US-0140353.
PR 23-JUN-1999; 99US-0140354.
PR 24-JUN-1999; 99US-0140695.

PR 28-JUN-1999; 99US-0140823.
PR 29-JUN-1999; 99US-0140991.
PR 30-JUN-1999; 99US-0141287.
PR 01-JUL-1999; 99US-0141842.
PR 01-JUL-1999; 99US-0142154.
PR 02-JUL-1999; 99US-0142055.
PR 06-JUL-1999; 99US-0142390.
PR 08-JUL-1999; 99US-0142803.
PR 09-JUL-1999; 99US-0142920.
PR 12-JUL-1999; 99US-0142977.
PR 13-JUL-1999; 99US-0143542.
PR 14-JUL-1999; 99US-0143624.
PR 15-JUL-1999; 99US-0144005.
PR 16-JUL-1999; 99US-0144085.
PR 16-JUL-1999; 99US-0144086.
PR 19-JUL-1999; 99US-0144325.
PR 19-JUL-1999; 99US-0144331.
PR 19-JUL-1999; 99US-0144332.
PR 19-JUL-1999; 99US-0144333.
PR 19-JUL-1999; 99US-0144334.
PR 19-JUL-1999; 99US-0144335.
PR 20-JUL-1999; 99US-0144352.
PR 20-JUL-1999; 99US-0144632.
PR 21-JUL-1999; 99US-0144684.
PR 21-JUL-1999; 99US-0144814.
PR 21-JUL-1999; 99US-0145086.
PR 22-JUL-1999; 99US-0145088.
PR 22-JUL-1999; 99US-0145089.
PR 22-JUL-1999; 99US-0145087.
PR 22-JUL-1999; 99US-0145089.
PR 22-JUL-1999; 99US-0145192.
PR 23-JUL-1999; 99US-0145145.
PR 23-JUL-1999; 99US-0145218.
PR 26-JUL-1999; 99US-0145224.
PR 26-JUL-1999; 99US-0145226.
PR 27-JUL-1999; 99US-0145913.
PR 27-JUL-1999; 99US-0145918.
PR 27-JUL-1999; 99US-0145919.
PR 28-JUL-1999; 99US-0145951.
PR 02-AUG-1999; 99US-0146386.
PR 02-AUG-1999; 99US-0146388.
PR 02-AUG-1999; 99US-0146389.
PR 03-AUG-1999; 99US-0147038.
PR 04-AUG-1999; 99US-0147204.
PR 04-AUG-1999; 99US-0147302.
PR 05-AUG-1999; 99US-0147192.
PR 05-AUG-1999; 99US-0147260.
PR 06-AUG-1999; 99US-0147303.
PR 06-AUG-1999; 99US-0147416.
PR 09-AUG-1999; 99US-0147439.
PR 09-AUG-1999; 99US-0147935.
PR 10-AUG-1999; 99US-0148171.
PR 11-AUG-1999; 99US-0148319.
PR 12-AUG-1999; 99US-0148341.
PR 13-AUG-1999; 99US-0148565.
PR 13-AUG-1999; 99US-0148684.
PR 16-AUG-1999; 99US-0149368.
PR 17-AUG-1999; 99US-0149175.
PR 18-AUG-1999; 99US-0149426.
PR 20-AUG-1999; 99US-0149722.
PR 20-AUG-1999; 99US-0149723.
PR 20-AUG-1999; 99US-0149929.
PR 23-AUG-1999; 99US-0149929.
PR 23-AUG-1999; 99US-0149930.
PR 25-AUG-1999; 99US-0149930.
PR 26-AUG-1999; 99US-0150586.
PR 26-AUG-1999; 99US-0150884.
PR 27-AUG-1999; 99US-0151065.
PR 27-AUG-1999; 99US-0151066.
PR 27-AUG-1999; 99US-0151080.
PR 30-AUG-1999; 99US-0151303.
PR 31-AUG-1999; 99US-0151438.
PR 01-SEP-1999; 99US-0151930.
PR 07-SEP-1999; 99US-0152363.

PR 10-SEP-1999; 99US-0153070.
 PR 13-SEP-1999; 99US-0153758.
 PR 15-SEP-1999; 99US-0154018.
 PR 16-SEP-1999; 99US-0154039.
 PR 20-SEP-1999; 99US-0154779.
 PR 22-SEP-1999; 99US-0155139.
 PR 23-SEP-1999; 99US-0155486.
 PR 24-SEP-1999; 99US-0156559.
 PR 28-SEP-1999; 99US-0156458.
 PR 29-SEP-1999; 99US-0156596.
 PR 04-OCT-1999; 99US-0157117.
 PR 05-OCT-1999; 99US-0157753.
 PR 06-OCT-1999; 99US-0157865.
 PR 07-OCT-1999; 99US-0158029.
 PR 08-OCT-1999; 99US-0158232.
 PR 12-OCT-1999; 99US-0158369.
 PR 13-OCT-1999; 99US-0159293.
 PR 13-OCT-1999; 99US-0159294.
 PR 13-OCT-1999; 99US-0159295.
 PR 14-OCT-1999; 99US-0159329.
 PR 14-OCT-1999; 99US-0159330.
 PR 14-OCT-1999; 99US-0159331.
 PR 14-OCT-1999; 99US-0159637.
 PR 14-OCT-1999; 99US-0159638.
 PR 18-OCT-1999; 99US-0159584.
 PR 21-OCT-1999; 99US-0160741.
 PR 21-OCT-1999; 99US-0160767.
 PR 21-OCT-1999; 99US-0160770.
 PR 21-OCT-1999; 99US-0160814.
 PR 21-OCT-1999; 99US-0160815.
 PR 22-OCT-1999; 99US-0160980.
 PR 22-OCT-1999; 99US-0160981.
 PR 22-OCT-1999; 99US-0160989.
 PR 25-OCT-1999; 99US-0161404.
 PR 25-OCT-1999; 99US-0161405.
 PR 25-OCT-1999; 99US-0161406.
 PR 26-OCT-1999; 99US-0161359.
 PR 26-OCT-1999; 99US-0161360.
 PR 26-OCT-1999; 99US-0161361.
 PR 28-OCT-1999; 99US-0161920.
 PR 28-OCT-1999; 99US-0161992.
 PR 28-OCT-1999; 99US-0161993.
 PR 29-OCT-1999; 99US-0162142.

Query Match 26.6%; Score 712; DB 21; Length 462;
 Best local Similarity 32.5%; Pred. No. 2.5e-54;
 Matches 164; Conservative 93; Mismatches 196; Indels 52; Gaps 8;

QY 12 LLLPLSLLLPLLLKRRN--RKTFRNLPCKSGMPFLGFTIGYLLKPYTATTLGDPMQ 69
 DB 2 MMILGLVLIIVICLCTALLRNNOMRYSKGLPPGTGMPFGEETFEFLKGP----DFMK 56
 QY 70 QHVSXGKIYRSNLFGEPTIVSADAGLNRFLLONEGRFLFOSYPRISIGILGKWSMLYV 129
 DB 57 NQRLRIGSFEKSHLGCPTIVSMDAELNRIILMNESGLVAGYPOSMDLIGTCNIAVH 116
 QY 130 GDMHRDMRSISLNFSLHARLRLTLKDYERHTLFLVDSWOONSIFSADDEAKKFTFNMA 189
 DB 117 GFSHRLMRGSLSLISPTMAKHLLPKIDPFMRNYLCGWDDLETVDIQEKTKHNAF--LSS 175
 QY 190 KHTMSMDPGEETEQQLKEVYPMKGVASAPNLPGTAHYKALOSRATILKFTIRKMEER 249
 DB 176 LLOIAETLKKRPEVEEYTFEFLVGTLSVPIDIPGTNYRSYGVARNNIDRLTLELMQER 235
 QY 250 KLDIKEDEEVEYKTEDEAEMSKSDHVRKQRTDDDLGWLKHSN----LSTEQILDLI 305
 DB 236 K-----ESGETFT-----DMLGYLMKKEDNRYLLTDKEIRDOY 268
 QY 306 LSLFAGHETSSVAIALAIFFLQACPAVELEREHLEIARAKKEGESEIMNDYDKMD 365
 DB 269 VTILYSGYEVTSMTMAKLYLHDHPRALBELREHAIIRKRKP--DEPLTLDIDIKSMK 326

QY 366 FTQCVINETLRIGNVVRFLHRRKALADVRYKGTIDISGMKVLVISAHLNDSRYDQPNLF 425
 DB 327 FTFAVIFETSRLATIVNGVLRKRTFDLELNGYLIPKGRNRYVYFREINVTSLYEDPMIF 386
 QY 426 NPMWQOQNNAGSSSGSFSFTWGNNNYMPFGGPRLCAGSELAKEMAVFIIHLYLKFNW 485
 DB 387 NPMWRMKSLESKS-----YFLFPGGVRILCPKEKELGISEVSSFLHYFVTKRW 435
 QY 486 ELAEDDQPFAPFVDFPNGLPIRVS 510
 DB 436 EENGEDKLMPFPRVSAPKGYHLKCS 460

Search completed: June 10, 2003, 21:57:23
 Job time : 46 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: June 10, 2003, 21:56:33 ; Search time 19 Seconds

(without alignments)
794.419 Million cell updates/sec

Title: US-09-502-426b-2

Perfect score: 2681 1 MFETHTLLPLLLPLSLLS.....FAFVDFPFGILPIRVSRL 513

Sequence: BLOSUM62

Scoring table: Gapop 10.0 , Gapext 0.5

Searched: 262574 seqs, 29422922 residues

Total number of hits satisfying chosen parameters: 262574

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

1: /cgn2_6/prodata/1/1aa/5A.COMB.pep:*
2: /cgn2_6/prodata/1/1aa/5B.COMB.pep:*
3: /cgn2_6/prodata/1/1aa/6A.COMB.pep:*
4: /cgn2_6/prodata/1/1aa/6B.COMB.pep:*
5: /cgn2_6/prodata/1/1aa/PCTUS.COMB.pep:*
6: /cgn2_6/prodata/1/1aa/Backfile1.pep:*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	978.5	36.5	472	2	US-08-622-166A-2
2	978.5	36.5	472	2	US-08-622-166A-4
3	503	18.8	492	3	US-08-724-466B-2
4	503	18.8	492	4	US-08-882-164D-2
5	461	17.2	497	4	US-08-882-164D-32
6	439	16.4	497	3	US-08-724-466B-4
7	439	16.4	497	4	US-08-882-164D-4
8	302.5	11.3	529	4	US-08-915-192A-405
9	289	10.8	511	4	US-08-991-677-4
10	279.5	10.4	513	3	US-08-948-564-6
11	273	10.2	504	4	US-09-144-367-2
12	271	10.1	504	4	US-09-126-420A-18
13	268.5	10.0	502	4	US-09-499-302A-2
14	268	10.0	467	4	US-09-126-420A-17
15	264	9.8	520	4	US-09-126-420A-26
16	264	9.8	520	2	US-09-091-432-2
17	263.5	9.8	504	1	US-08-457-274A-5
18	263.5	9.8	504	1	US-08-457-274A-25
19	257	9.6	476	1	PCT-US95-05758-25
20	256.5	9.6	490	1	US-08-313-075A-10
21	256.5	9.6	490	1	US-08-201-118-7
22	256.5	9.6	490	5	PCT-US95-05744-7
23	254.5	9.5	509	4	US-09-499-302A-8
24	251	9.4	504	4	US-09-499-302A-6
25	250.5	9.3	504	4	US-09-126-420A-23
26	245	9.1	495	1	US-08-532-065B-2
27	244.5	9.1	576	3	US-08-948-564-16

28	242.5	9.0	502	4	US-09-499-302A-4	Sequence 4, Appl 1
29	241.5	9.0	531	4	US-09-380-420C-2	Sequence 2, Appl 1
30	241	9.0	490	4	US-09-126-420A-16	Sequence 16, Appl 1
31	241	9.0	510	3	US-08-606-505B-66	Sequence 66, Appl 1
32	241	9.0	510	4	US-09-616-990-65	Sequence 65, Appl 1
33	240.5	9.0	509	3	US-08-948-564-18	Sequence 18, Appl 1
34	239	8.9	496	1	US-08-313-075A-50	Sequence 50, Appl 1
35	236	8.8	521	3	US-08-948-564-14	Sequence 14, Appl 1
36	234.5	8.7	430	1	US-08-201-118-1	Sequence 1, Appl 1
37	234.5	8.7	490	2	US-08-238-821B-1	Sequence 1, Appl 1
38	234.5	8.7	490	5	PCT-US95-05744-1	Sequence 1, Appl 1
39	233.5	8.7	508	4	US-08-991-677-2	Sequence 2, Appl 1
40	232.5	8.7	504	4	US-09-499-302A-9	Sequence 9, Appl 1
41	231	8.6	506	1	US-08-313-075A-38	Sequence 38, Appl 1
42	231	8.6	506	3	US-08-606-505B-65	Sequence 65, Appl 1
43	231	8.6	506	4	US-09-616-990-65	Sequence 65, Appl 1
44	225.5	8.4	523	4	US-09-302-620B-95	Sequence 95, Appl 1
45	225	8.4	471	4	US-09-126-420A-20	Sequence 20, Appl 1

ALIGNMENTS

RESULT 1

US-08-622-166A-2

Sequence 2, Application US/08622166A

Patent No. 5953545

GENERAL INFORMATION:

APPLICANT: KONCH, CSABA

APPLICANT: MATHUR, JAIDEEP

APPLICANT: SZEKERES, MIKLOS

APPLICANT: ALTMANN, THOMAS

TITLE OF INVENTION: NUCLEIC ACID MOLECULES ENCODING CYTOCHROME

NUMBER OF SEQUENCES: 4

ADDRESS: Birch, Stewart, Kolasch & Birch, LLP

STREET: P.O. Box 747

CITY: Falls Church

STATE: Virginia

COUNTRY: USA

ZIP: 22040-0747

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk

OPERATING SYSTEM: IBM PC compatible

SOFTWARE: Patent Release #1.0, Version #1.30 (EPO)

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/08/622,166A

FILING DATE: 27-MAR-1996

CLASSIFICATION: 800

ATTORNEY/AGENT INFORMATION:

NAME: SVENSSON, LEONARD R.

REGISTRATION NUMBER: 30,330

REFERENCE/DOCKET NUMBER: 0147-0153P

TELECOMMUNICATION INFORMATION:

TELEPHONE: (703) 205-8000

TELEFAX: (703) 205-8050

TELEX: 248345

INFORMATION FOR SEQ ID NO: 2:

SEQUENCE CHARACTERISTICS:

LENGTH: 472 amino acids

TYPE: amino acid

TOPOLOGY: linear

MOLECULE TYPE: protein

US-08-622-166A-2

Query Match 36.5%; Score 978.5; DB 2; Length 472;

Best Local Similarity 41.9%; Pred. No. 5e-69; Mismatches 171; Indels 41; Gaps 8;

Matches 210; Conservative 79; Mismatches 171; Indels 41; Gaps 8;

12 LLLPLSLILFLILKRRNRKTRFNLPPKSGWPLGFTIGYIKPYATATLADFMQOH 71.

[illegible]

Query Match	18.8%	Score 503;	DB 3;	Length 492;
Best Local Similarity	27.6%	Pred. NO. 1.5e-41;		
Matches 144;	Conservative 95;	Mismatches 216;	Indels 66;	Gaps 14;

RESULT 4
US-08-882-164D-2

Patent No. 6306692A:
GENERAL INFORMATION:
APPLICANT: Petkovich, P. Martin, Whit
APPLICANT: Beckett, Barbara R., Jones
TITLE OF INVENTION: Retinoid Metabol
NUMBER OF SEQUENCES: 43
CORRESPONDENCE ADDRESS:
ADDRESSEE: Blake, Cassels & Graydon
STREET: Box 25, Commerce Court West
CITY: Toronto
STATE: Ontario
COUNTRY: Canada
ZIP: M5L 1A9
COMPUTER READABLE FORM:

Query Match	18.8%	Score 503;	DB 4;	Length 492;
Best Local Similarity	27.6%	Pred. No. 1.5e+11;		
Matches 144;	Conservative 95;	Mismatches 216;	Indels 66;	Gaps 14;

```

QY HTLLPLLLPSLLSLFL-----ILKRNKRKTPNLPKGSGRPLGELIGVLP 58
Db 4 YTLMTVECTIYVPLVLLAAKRLMEMLMIRRVDPNCSPLPBGMLPLFGELIOLI-- 61
QY 59 Y7ATTLGDEMOQHVSKYKIVRSNLPGEPTIVSADAGINREPLONEGRLEFCSYPSRIGG 118
Db 62 ---LORRKEFLRMKROKYGCITYKTHLFGNPTVAVMGADVWROLLEGHKLVSVQNPASVRT 118
QY 119 ILKWSMLVYVGDMMHD-----MRSISLNFISHALRLLILKQVBRHLPVLDSE--QONS 172
Db 119 ILSDDLNSVNGVQHKKKAKIMRAFSRDALSH-----YIPVIOOEYKSAIOEWLOKDS 172
QY 173 IFSADQAKFTFNLMAKHMSMDGEEET--EOLKKEYVTFMGVVSAPLNLPGTAYHK 230
Db 173 CVLVYEMKMKMERIMRILLFEEPOJKTDEQELVEAFEMIKNPLSLPIDVPFSGLYR 232
QY 231 ALOSRAATILKFERKMEERKLDIKEDOEDEEVKTEDAEKSKDHRKQRTDDLLGW 250
Db 233 GLRAR----NFHSIKTEENIRKKIIDDENNQOKMLDOLL-----IENSRSDE----- 279
QY 291 LKHSNSTBOILLDLSLFPAGHETSVAVALAIFLOACPAAVEELREHLELTARAKE 350
Db 280 ----PESLOAMEKAEATELLFGSHETTASTATSLVMFLGNTVEVQKREE----YQKKE 331
QY 351 LG-----ESLNMDDYKKKDFQOCYVINEPLRLGNYRPLHKKALDVKRYKGTIDPSGKVL 406
Db 332 MGYTPRGKLSMELLDOUKYTCGVAKETELRLINPVRGGFRVALTFEPLNGYQIPIKGMNVI 391
QY 407 PVLSAVHLNDSRYDOPNLFPNPRMOOONNGASSGSGSEFTWGNVNYRFGGGLPGSGSE 466
Db 392 YSICDTHVADVAPFKKEEFOREN-----SKGLEGSGSF--NYIPFGGSMCYGKE 442
QY 467 LAKLEAVNTIHLVLKFNWELAEDDO-----PFAAPVDP 502
Db 443 FAKVLLKTFVLVELQHCNMWILSNGPRTKTPPTIYVNDLP 483

```

RESULT 5
US-08-882-164D-32
; Sequence 32, Application. US/08882164D
; Patent No. 6306624
; GENERAL INFORMATION:

```

APPLICANT: Petkovich, P. Martin, White, Jay A.,
APPLICANT: Beckett, Barbara R., Jones, Glenville
TITLE OF INVENTION: Retinoid Metabolizing Protein
NUMBER OF SEQUENCES: 43
CORRESPONDENCE ADDRESSES:
ADDRESS: Blake, Cassels & Graydon
STREET: Box 25, Commerce Court West
CITY: Toronto
STATE: Ontario
COUNTRY: Canada
ZIP: M5L 1A9
COMPUTER READABLE FORM:
MEDIUM TYPE: Diskette, 3 1/2 inch, 1.4 Mb storage
COMPUTER: COMPAQ, IBM PC compatible
OPERATING SYSTEM: MS-DOS 5.1
SOFTWARE: WORD PERFECT
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/882,164D
FILING DATE: June 25, 1997
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 08/667,546
FILING DATE: June 21, 1996
APPLICATION NUMBER: 08/724,466
FILING DATE: October 1, 1996
ATTORNEY/AGENT INFORMATION:
NAME: Hunt, John C.
REGISTRATION NUMBER: 36,424
REFERENCE/DOCKET NUMBER: 50767/00010
TELECOMMUNICATION INFORMATION:
TELEPHONE: (416) 863-4344
TELEFAX: (416) 863-2653
INFORMATION FOR SEQ ID NO: 32:
SEQUENCE CHARACTERISTICS:
LENGTH: 497 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
US-08-882-164D-32

Query Match          17.2%; Score 461; DB 4; Length 497;
Best Local Similarity 28.4%; Pred. No. 2.4e-37;
Matches 148; Conservative 78; Mismatches 231; Indels 64; Gaps 14;

QY 10 LPALLPSL-----LSLLFLILK-----RRNRKTRFNLPPKSGMPFLGTYGLK 57
DB 3 LPALLASALCTFVLPPLFLAALKIMDLVCVSGRSCALPLPPTMGFPFFGTYL---- 58
QY 58 PYATATIGDFMOQHVSYKGIYRSNLGEPTIVSADAGLNFTIIONEGRLFECSPRSIG 117
DB 59 -QWVLOKRRKFLQMKRRKTYGTYKTHLFGRPVYRWGADNVRRIILGHRIVSVHMPASVR 117
QY 118 GILGKMSMLVVGDMHRMSISLNFSLHARLFTILKDVERRHLFLVDSWQO--NSIFS 175
DB 118 TIIGAGCLSNLHDSHOKRKVIMQAFSREALQCYLV-IAEYSSCLEQMLSCGEGGL 176
QY 176 AODEAKFTTNLAKKHIMSDPGF---EETEDLKREYVTFMKGVSAPIMLPCTAYHKA 231
DB 177 VYEVEVKRLMFIRIMRILGCEPAGGEGDEQOLVEAFEEETRLNLSPLIDVPSGLYRG 236
QY 232 LOSRATLTKFERKMEER--RKLDIKEED---QEEEEKTEDEAMSDHVRKRTDD 286
DB 237 VKARNLHARIEENIRAKIRRLQATEPDDGCKDLQLLIHSW-----RGERLDMQ- 288
QY 287 LGWVLKHSNLSTEOIDLILSLFAGHETSVAIALAIFLQACPKAVEELREHLEIAR 346
DB 289 ---ALNOS---STE-----LFGGHETASATSLITVLGLYPHYLVQVREIRKSKGL 335
QY 347 AKKELESEELNMDYKMDFTQCYINETTLRGVNVRELRKALKDVYKYGYDIPSGMKVL 406
DB 336 LCKSDNKNLDMETLEQLKYIGVYIKETRLNPNPVPGFRVALKTFELNGLQIPKGNVI 395
QY 407 PVTSANVHDSNRDQPLFNPMMQOONNCASSSGSSTWGNNTYPPFGGGLPCGSE 466

```

```

DB 396 YSICDTHVDIIFTNKEEFNDPRTVPHPEASRFS-----FIFGGGLRSCVCKE 446
QY 467 LAKLEMAVEFIHLVLFKNMELAE-----DDQFAFPYDFP 502
DB 447 FAKILKIFVYELARHCMDQLNGPPTKTSPTVYPVNDLP 487

RESULT 6
US-08-724-466B-4
Sequence 4, Application US/08724466B
Patent No. 6063606
GENERAL INFORMATION:
APPLICANT: Petkovich, P. Martin, White, Jay A.,
APPLICANT: Beckett, Barbara R., Jones, Glenville
TITLE OF INVENTION: Retinoid Metabolizing Protein
NUMBER OF SEQUENCES: 30
CORRESPONDENCE ADDRESS:
ADDRESS: Blake, Cassels & Graydon
STREET: Box 25, Commerce Court West
CITY: Toronto
STATE: Ontario
COUNTRY: Canada
ZIP: M5L 1A9
COMPUTER READABLE FORM:
MEDIUM TYPE: Diskette, 3 1/2 inch, 1.4 Mb storage
COMPUTER: COMPAQ, IBM PC compatible
OPERATING SYSTEM: MS-DOS 5.1
SOFTWARE: WORD PERFECT
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/724,466B
FILING DATE: October 1, 1996
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 08/667,546
FILING DATE: June 21, 1996
ATTORNEY/AGENT INFORMATION:
NAME: Hunt, John C.
REGISTRATION NUMBER: 36,424
REFERENCE/DOCKET NUMBER: 50767/00004
TELECOMMUNICATION INFORMATION:
TELEPHONE: (416) 863-4344
TELEFAX: (416) 863-2653
INFORMATION FOR SEQ ID NO: 4:
SEQUENCE CHARACTERISTICS:
LENGTH: 497 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
US-08-724-466B-4

Query Match          16.4%; Score 439; DB 3; Length 497;
Best Local Similarity 27.0%; Pred. No. 3.7e-35;
Matches 142; Conservative 85; Mismatches 226; Indels 72; Gaps 15;

QY 10 LPALLPSL-----LSLLFLILK-----RRNRKTRFNLPPKSGMPFLGTYGLK 57
DB 3 LPALLASALCTFVLPPLFLAALKIMDLVCVSGRSCALPLPPTMGFPFFGTYL---- 58
QY 58 PYATATIGDFMOQHVSYKGIYRSNLGEPTIVSADAGLNFTIIONEGRLFECSPRSIG 117
DB 59 -QWVLOKRRKFLQMKRRKTYGTYKTHLFGRPVYRWGADNVRRIILGHRIVSVHMPASVR 117
QY 118 GILGKMSMLVVGDMHRMSISLNFSLHARL--TILKDVERRHLFLVDSWQO--NS 172
DB 118 TIIGAGCLSNLHDSHOKRKVIMQAFSREALQCYLV-IAEYSSCLEQMLSCGEGGL 176
QY 173 IFSQODEAKFTTNLAKKHIMSDP-----QEEETEDLKREYVTFMKGVSAPIMLPETA 227
DB 174 GLVLYPEVKRLMFIRIMRILGCEPQLAGDSDS-ODLVEAFEEETRLNLSPLIDVPSG 232
QY 228 YHKLQSRATLTKFERKMEERKLDIKEED---QEEEEKTEDEAMSDHVRKORT 282
DB 233 LYRGMKARNLHAIIEONIRAKITGLRASEAGGCKDLQLLIHSW-----RGERL 285
QY 283 DDLGLGWVLKHSNLSTEOIDLILSLFAGHETSVAIALAIFLQACPKAVEELREHLEIAR 342

```

Db 286 DMO---ALKOS---STE-----LIFGHEHTASATSLITVGLYHVLQVREELK 331
QY 343 EIAARAKKEGESELMNDYKMDPTQCVINETRLGNAVVELHRAKALDVKRYGYDIPSG 402
Db 332 SKGLLCKSNODNKLDMELIQLKYGICVIKETRLNPPVGGFRVALKTFELNGYQIPKG 391
QY 403 WKVLPYISAVHLNDSRYDOPNLFNPMWOOQNNGASSSGSFTWGNMMPGGGPRLC 462
Db 392 WNYIYSICDTHDVAEIFTKNEEPNDRFSAPHEDASRFS-----FIFPGGLRSC 442
QY 463 AGSELAKLEMAVETIHLVLFKNWELAE-----DDQFAFPFVDFP 502
Db 443 VGKEFAKILIKITFTVELARHCDWLNGPPTMKTSPYVVDNLP 487

RESULT 7

US-08-882-164D-4
; Sequence 4, Application US/08882164D
; Patent No. 6306624
; GENERAL INFORMATION:
; APPLICANT: Petkovich, P. Martin, White, Jay A.,
; APPLICANT: Beckett, Barbara R., Jones, Glenville
; TITLE OF INVENTION: Retinoid Metabolizing Protein
; NUMBER OF SEQUENCES: 43
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Blake, Cassels & Graydon
; STREET: Box 25, Commerce Court West
; CITY: Toronto
; STATE: Ontario
; COUNTRY: Canada
; ZIP: M5L 1A9
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette, 3 1/2 inch, 1.4 Mb storage
; COMPUTER: COMPAQ, IBM PC compatible
; OPERATING SYSTEM: MS-DOS 5.1
; SOFTWARE: WORD PERFECT
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/882,164D
; FILING DATE: June 25, 1997
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/667,546
; FILING DATE: June 21, 1996
; APPLICATION NUMBER: 08/724,466
; FILING DATE: October 1, 1996
; ATTORNEY/AGENT INFORMATION:
; NAME: Hunt, John C.
; REGISTRATION NUMBER: 36,424
; REFERENCE/DOCKET NUMBER: 50767/00010
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (416) 863-4344
; TELEFAX: (416) 863-2653
; INFORMATION FOR SEQ ID NO: 4:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 497 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; US-08-882-164D-4

Query Match 16.4%; Score 439; DB 4; Length 497;
Best Local Similarity 27.0%; Pred. No. 3.7e-35;
Matches 142; Conservative 85; Mismatches 226; Indels 72; Gaps 15;
QY 10 LPULLPSL-----LSLLFLILK-----RRNRKTRFNLPPKSGMPLGEGTIGYLK 57
Db 3 LPALLASLCFTVPLPLFLAIIKIMLDYCVSGRDRSCALPLPCTMGFPFEGEL-----58
QY 58 PYTATTLGDFMQOAHKSKYGIYSNLGEPPIYASADALNFIIONBRLTECSPPSIG 117
Db 59 -QWVLRKRFIQMRKRTGFIYKTHLFRPTVRWAGADNVRRIILLGDRLVSVHRPASVR 117
QY 118 GILGKWSMLVVLGDMHRMRSISINFLSHARLR--TILLADVERHTLFLVDSMOQ--NS 172

Db 118 TILSGGSLNLDHSSHQRRKVIKRAFSREALRCYVVIIEVGSS-----LEQWLSCGER 173
QY 173 IFSAOEAKKFTFNLMAKHHMSMDP-----GEBETDOLKEVYTFPMKVSAALNPGTA 227
Db 174 GLVYEPVKRLMRIMRILLEGCEPOLAGGDSB--QOLVAEPEKMTNLFSLPIDVFGS 232
QY 228 YHVALDSRATILKFERKMEERLDIKEED-----QEEBEKTEDEAKMSKSDHVRKQRT 282
Db 233 LYRGKARNLIHARIQINIRAKICGLRASAAGGCCDAQQLLEHSM-----RGERL 285
QY 283 DDOLLGMYLKHSLNTEQIIDLILLSLFAGHESSVALATLFFLOACRANAELEBEH 342
Db 286 DMO---ALKOS---STE-----LIFGHEHTASATSLITVGLYHVLQVREELK 331
QY 343 EIAARAKKEGESELMNDYKMDPTQCVINETRLGNAVVELHRAKALDVKRYGYDIPSG 402
Db 332 SKGLLCKSNODNKLDMELIQLKYGICVIKETRLNPPVGGFRVALKTFELNGYQIPKG 391
QY 403 WKVLPYISAVHLNDSRYDOPNLFNPMWOOQNNGASSSGSFTWGNMMPGGGPRLC 462
Db 392 WNYIYSICDTHDVAEIFTKNEEPNDRFSAPHEDASRFS-----FIFPGGLRSC 442
QY 463 AGSELAKLEMAVETIHLVLFKNWELAE-----DDQFAFPFVDFP 502
Db 443 VGKEFAKILIKITFTVELARHCDWLNGPPTMKTSPYVVDNLP 487

RESULT 8

US-09-615-192A-405
; Sequence 405, Application US/09615192A
; Patent No. 6410718
; GENERAL INFORMATION:
; APPLICANT: Bloksberg, Leonard N.
; APPLICANT: Havukkala, Ilkka
; TITLE OF INVENTION: Materials and Methods for the
; FILE REFERENCE: 11000.1003c4U
; CURRENT APPLICATION NUMBER: US/09/615,192A
; CURRENT FILING DATE: 2000-07-12
; PRIOR APPLICATION NUMBER: US 08/975,316
; PRIOR FILING DATE: 1997-11-21
; PRIOR APPLICATION NUMBER: US 08/713,000
; PRIOR FILING DATE: 1996-09-11
; PRIOR APPLICATION NUMBER: US 09/169,769
; PRIOR FILING DATE: 1998-10-09
; NUMBER OF SEQ ID NOS: 405
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO: 405
; LENGTH: 529
; TYPE: PRT
; ORGANISM: Eucalyptus grandis
; US-09-615-192A-405

Query Match 11.3%; Score 302.5; DB 4; Length 529;
Best Local Similarity 21.9%; Pred. No. 1.7e-21;
Matches 117; Conservative 104; Mismatches 204; Indels 109; Gaps 18;
QY 11 PULLPSLILLPLILKRRNRKTRFNLPPKSGMPLGEGTIGYLKPYATTLGDFMQO 70
Db 22 PFTLLSVPLLLFLGLVARLRKRPF--PGPRLPLVIGNML-----MMGEILHR 70
QY 71 HVS-----KYGKIYSNLGEPPIYASADAGLRFIIONBRLTECSPPSIG 116
Db 71 GLASLAKTYGGIFHLRMOFLIMNAVSSDVARQVLOVHDGIFSNRPATIAISYLYDAD 130
QY 117 GILGKWSMLVVLGDMHRMRSIS--LNFLSHARLTILKADVERHTLFLVDSMOQNSIFS 175
Db 131 -----MAFAHYGPFQMRLCYMKLFSRKR-----AESWE-----S 162
QY 176 AODEAK-----KTFNLMAKHHMSMDPGEETEOLKKEVYTFMK-----214
Db 163 VRDEVDIMVRIVASGEGTAVNIGELVLELTDITVRAAFGTSSTEG--ODEFISILQESK 221

```

QY 215 --GVASAPLNLPATV-----HKAQSRAITLIERKME--ERKLIDKEDEEE 262
| : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 222 LFGAFNADFLPYLSWIDPGLTARLVKAROSLDGFIHIDHDKRNKTSGGGD 281
QY 263 VKTE--DEAEKMSGDVHRKORTDDLLGWLKHS--NISTEQIDLILSLNAGHTSSVA 319
| : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 282 VDTMDVDDLAFYSDEAVNESD-----LQNSIRLRDNKIKAIMVMGGETVASA 335
QY 320 IATAIFLQACPKAVELEREHLEIARAKKEGSELMDYKKMDFQCVINETLRGN 379
| : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 336 IEMAMELMRSPEDLKKVQOELADYVGIDRVEES----DEKITYIAKCKLKLRLHP 390
QY 380 VVFLHRAKALKDVYKGYDIPSGWKVLPVISAHLNDSRYDQPLNFWMOQONGASS 439
| : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 391 PIPLLHETAEADAVISGRIPARSRYMINAMWIGRDCSWTEPDKFKPSRLE----- 443
QY 440 SGSGSFSTWGNVNPFGGPRLCAGSELAKLEMAVFTHLVKLNWELAEEDOP 493
| : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 444 SGMPDYKGSNFEPIFGSGRRSCPMQGLYALDMAVAHLHCFTWELPDGMRP 497

```

RESULT 9

```

US-08-991-677-4
; Sequence 4, Application US/08991677A
; Patent No. 6252135
; GENERAL INFORMATION:
; APPLICANT: Chianq, Vincent L
; APPLICANT: Caraway, Daniel T
; APPLICANT: Smeltzer, Richard H
; TITLE OF INVENTION: Production of Strylingyl Lignin in Gymnosperms
; FILE REFERENCE: 50617
; CURRENT APPLICATION NUMBER: US/08/991,677A
; CURRENT FILING DATE: 1997-12-16
; EARLIER APPLICATION NUMBER: US 60/033,381
; EARLIER FILING DATE: 1996-12-16
; NUMBER OF SEQ ID NOS: 11
; SOFTWARE: Patentn Ver. 2.0
; SEQ ID NO 4
; LENGTH: 511
; TYPE: PRT
; ORGANISM: Liquidambar styraciflua
US-08-991-677-4

```

```

Query Match 10.8%; Score 289; DB 4; Length 511;
Best Local Similarity 22.0%; Pred. No. 3.6e-20;
Matches 121; Conservative 95; Mismatches 196; Indels 138; Gaps 21;

```

```

QY 6 HHTLLP-LLLPSLLSLFLILKRRNRKTRFNLPPKSGWPLGEGTIGLYATTL 64
| : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 6 HEALQPLPMTLFFIIPILLGLVSR--LRQRLPYPGPKGLPYIGMWL--MMDQLTRGL 62
QY 65 GDFMOQVSKYKRIYRSLFGEPTIVSADAGLNRFIIONEGRLPE-----CSYPR 114
| : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 63 AKLAKQ---YGGFLHAKGFLHVAVSTDMARQVQOVQDNISNRPATIAISLYYDR 118
QY 115 SIGGILKWSMLVLVGDHMRMSIS--LNFLSHARLRTILLKDVERTLLFVLSMOONS 173
| : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 119 AD-----MAFAHYGPFMRQMKLCVMLFSRKR-----AESME----- 151
QY 174 FSNODE-----AKFTFNLAKHIMSDPGE-----ETE 203
| : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 152 -SVRDEVDAVRVANSIGSTVIGELVFALTKNITYRAFGTISHDEDEFAVILQEF 210
QY 204 QL-----KREYVTFMGV---VSAPLNLPGTAYHKALOSRAITLIERKMEERLIDKE 255
| : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 211 QLGAFVIADEFIPLKLVNPGGINVRLN-----KAGALDGFIDKIIDH---IQK 257
QY 256 EDGEEVEVTE-----DEAEKMSGDVHRKORTDDLLGWLKHSNLSSTEQIIDLI 305
| : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 258 GSKNSEVDVDMVDLLAFYGEAKVSESDD-----LQNSIKLTKDNIKAI 303
QY 306 LSLFLAGHTSSVAIALAIFLQACPKAVELEREHLEIARAKKEGSELMDYKKMD 365

```

```

Db 304 MDVNGGETVASALEMAMTELMKSPEDLKKVQOELAVVVGIDRVEEK-----DEKLT 358
QY 366 FQOAGINFLRGNVVRFLHRAKALKDVYKGYDIPSGWKVLPVISAHLNDSRYDQPLN 425
| : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 359 YLKCYLEVLRHAPPIPLHETAEADAEVGGIYIIRAKSVYMINACAIKGDKSNADPDR 418
QY 426 NFWMOQONGASSSGSFSTWGN--YMPFGGPRLCAGSELAKLEMAVFTHLVKLF 483
| : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 419 RPSRLK-----DGVPDK--GNMFEPIFGSGRRSCPMQGLYALLETVAHLHCF 469
QY 484 NWELEADOP 493
| : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 470 TWELPDGMRP 479

```

RESULT 10

```

US-08-948-564-6
; Sequence 6, Application US/08948564
; Patent No. 6121512
; GENERAL INFORMATION:
; APPLICANT: Siminszky, Balazs
; APPLICANT: Dewey, Ralph E.
; APPLICANT: Corbin, Frederick T.
; TITLE OF INVENTION: No. 6121512el Cytochrome P-450 Constructs and
; TITLE OF INVENTION: Methods of Producing Herbicide-Resistant Transgenic Plant
; NUMBER OF SEQUENCES: 23
; CORRESPONDENCE ADDRESS:
; ADDRESS: Virginia C. Bennett
; STREET: PO Box 37428
; CITY: Raleigh
; STATE: No. 6121512th Carolina
; COUNTRY: USA
; ZIP: 27627
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/948,564
; FILING DATE:
; CLASSIFICATION: 800
; ATTORNEY/AGENT INFORMATION:
; NAME: Bennett, Virginia C.
; REGISTRATION NUMBER: 37,092
; REFERENCE/DOCKET NUMBER: 5051-409
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 919-854-1400
; TELEFAX: 919-854-1401
; INFORMATION FOR SEQ ID NO: 6:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 513 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
US-08-948-564-6

```

```

Query Match 10.4%; Score 279.5; DB 3; Length 513;
Best Local Similarity 22.9%; Pred. No. 3.2e-19;
Matches 118; Conservative 82; Mismatches 216; Indels 99; Gaps 17;

```

```

QY 15 LPSLLSLFLILKRRNRKTRFNLPPKSGWPLGEGTIGLY---KPYTATLADGFMQOH 71
| : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 15 LAFISGLIF--FLKQSKSKRPNLPFGPMPIYGNLFQVARSCKPFF-----EYVDV 67
QY 72 VSKYGIKYSNLFGEPTIVSADAGLNRFIIONEGRLFCYSYPR----- 114
| : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 68 RLKYSIFTLKMGRTMILLDADKLVHBMLOKATVATRPENPTFIIFSENKFTVNA 127
QY 115 SIGGILKWSMLVLVGDHMRMSISLNFLSHARLRTILLKDVERTLLFVLSMOONS 168
| : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 128 TYGPV---WKSLL-----RRNVQNMILSSTRKE--FRSVNRDNAMDKLINRLKDAE 173

```



```

QY 169 OONSIFSAQDCAKKTFFNIMKHNISMDOGEEETQKKREYTFMKGVSA----- 219
D 174 KNGVWVWKARRRFAVFCILVAMCFLGEMDEETVERIDP-----VMSKLTITDPRIDYL 229
QY 220 PLNLP--GTAAYKALOSRATILKFTERKMEERKIDKEBDQEBEEVKTEDAEMSKSDHY 277
D 230 PILSPFESKOKKALEFVRQOEYFVLPIIEQRRALONG-----SDHT 273
QY 278 RKQRTDDLL-----GMYLKHNSLSTQDIDLLILSLFAGHETSVAIALAIFFLQACP 331
D 274 ATFSYLDTLPLDKVEG---KKSASDAELVSLCSEFLNGSDITLTAIVEMGIAQLIANP 330
QY 332 KAVELKREHELEIARAKKELGESLENMDYKMDFTQCIYNETLRIGLVNVRFLHRAKLD 391
D 331 NVQTKLYEE-----IKRTVEKKYDKEKDEKEMPRYLHNAVVELLKRHPPLTHFVLTHAVTE 384
QY 392 -VRKGYDIPESGMKVLPIYISAVHLDNSKRYDQNLNPNRWQOONNCASSGSSSESTWGN 450
D 385 PTLGVDYDIPIDANEVYTPAIABECPKMLNLEKFPDPERFISIGSEBADITG-----VTGV 439
QY 451 NYMPFGGPRILCAGSELKLEMAVYIHHVLKLENN 485
D 440 KMEFPGVGRILCPGLAMATVHHHLMAMARVQOEFEM 474

```

```

RESULT 11
US-09-144-367-2
: Sequence 2, Application US/09144367
: Patent No. 6432639
: GENERAL INFORMATION:
: APPLICANT: Lichter, Jay
: TITLE OF INVENTION: GENOTYPING OF HUMAN CYP3A4
: FILE REFERENCE: SEQ-12P
: CURRENT APPLICATION NUMBER: US/09/144,367
: PRIOR FILING DATE: 1998-08-31
: PRIOR FILING DATE: 1997-09-10
: NUMBER OF SEQ ID NOS: 58
: SOFTWARE: FastSeq for Windows Version 3.0
: SEQ ID NO 2
: LENGTH: 503
: TYPE: PRT
: ORGANISM: H. sapiens
: US-09-144-367-2

```

[illegible]

```

Dh      283 ETE-----SHKALSDLEIAOSIIFIIPAGYEITTSVLSEFIWELATHEDVQKQDE 334
QY      341 HETIAPAKKEIASESELNMDYKKMDFTQCVINETIRLGNAVVRFLHRAKLDVRYKGYDIP 4000
      :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: ::
Dh      335 -IDAVLPNK----APPTYDYLQMEYILMVVNETLRLPEIRAMLEREYCKKQVEINGMIF 3899
QY      401 SGMKVLPVLSAVHLDNSKRDQNLNPNPRMOOQNNNGASSSSGGSFSTWGNK-----YMPF 4555
      :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: ::
Dh      390 KGVNVVIRPSYALHHRPKYWTPEKEKLPERSKNN-----DNIDPYIYTPF 4355
QY      456 GGGPRCAGSELAKLEMAVEFIHILHKKENWELAEEDQFAFPVDFPNGLRIRVS 510
      ||| ||| ||| :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: ::
Dh      436 GSGPRRCIGTMRALNMKIALIRVLQNFSEPKCKETO-----IPLKIS 478

```

```

RESULT 12
US-09-126-420A-18
/ Sequence 18, Application US/09126420A
/ Patent No. 637673
/ GENERAL INFORMATION:
/ APPLICANT: BATARD, YANNICK
/ APPLICANT: ROBINEAU, TIBURCE
/ APPLICANT: DURST, FRANCIS
/ APPLICANT: WERCK-REICHART, DANIELE
/ APPLICANT: DIDIERJEAN, LUC
/ TITLE OF INVENTION: PURIFIED CYTOCHROME P450 CYP76A1 FROM HELIANTHUS
/ TITLE OF INVENTION: TUBEROSUS AND ITS APPLICATIONS AS BIOCATALYST IN
/ TITLE OF INVENTION: PARTICULAR FOR THE DEGRADATION OF ENVIRONMENTAL
/ TITLE OF INVENTION: POLLUTANTS AND FOR ALTERING THE RESISTANCE OF PLANTS S
/ FILE REFERENCE: 03715_0032
/ CURRENT APPLICATION NUMBER: US/09/126,420A
/ CURRENT FILING DATE: 1998-07-30
/ PRIOR APPLICATION NUMBER: 60/034,351
/ PRIOR FILING DATE: 1997-07-31
/ NUMBER OF SEQ ID NOS: 27
/ SOFTWARE: PatentIn Ver. 2.1
/ SEQ ID NO 18
/ LENGTH: 504
/ TYPE: PRF
/ ORGANISM: Solanum melongena
/ US-09-126-420A-18

```

Query Match	10.1%	Score 271;	DB 4;	Length 504;
Best Local Similarity	21.7%	Pred. No. 2,2e-18;		
Matches	111;	Conservative	95;	Mismatches 231; Indels 74; Gaps 17
QY	12	LLLLPSLLSLFLILKRRNRKRTFNLPRKSGMPFLGTTGY-LKPYATATLGDPMQ	70	
Db	13	IIILPA-----FILFESQKNTTKTSYPGPGLPIRGNNFELGTETPYKMAV---LRQ	63	
QY	71	HYSKYKIRSLSEEPITVYSADAGLNPFILQNEGRLE-----ECSTPRSTGGT	119	
Db	64	---KIGPVLMLKEGSTYTMVVOGTAAASELKRNMHDISPANRVIPDVNOAHSTYQGSIA--	118	
QY	120	LGMWSMLVVGDMHSDMSI--SLNFLSHARLFTILKVERHTFLVDSMOQNSIFSAOD	178	
Db	119	-----IAPYGFPMFQRRICITIEFYAKKISE--TEPYRKACVDNMLKTEKEANAER	170	
QY	179	---EAKKF---TNNLAKHIMSMDPGEETEDLKKEYVTFMK-----GVASAPLNLPG	225	
Db	171	GSIGIEYTRFVFLASNMGLNLSKDLADLSESESEFFIMAKRIINMSGIANSYDIFPE	230	
QY	226	TAYHAKLSRATILFPIERKKMERKLDIKEDDQEEVUKTDEKMSKSDHYRKQRPDD	285	
Db	231	LKKFDLQSLRRKMAWDGKAVEIMSFLKE--REBERKKGTCKRDLVDLVLEFQGGKD	288	
QY	286	LLGWVLKHSNSTEODILDLILSLFAGHETSSVAIALAIFLQACPRAVEIREBHETEA	345	
Db	289	-----EPRAKSEHEIKIFVLEMLAGETTTSSVEMALTELLNHPAMAKVTEISQAI	342	
QY	346	RAKKELGSEBELMWDDYKKMDPTQCYINETLALGVNVRFL-IRKALKVYRKYGIDIPSGWK	404	
Db	343	EPNRKFEOS-----DIENIPYMQAVLKSLSLAPRLPFLPRETIQOTKMGVAVDKDQ	397	

```

QY      405  VLEPVISAVHLDNSRYQOPLFNPMMMOQONNGASSGSGSPSTMGNNV--WPEFGGRLC 462
          | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db      398  VLVNAMAIGRDECDMDPMFKPEKFL-----GSKIDVGGHGLIFPAAGRRMC 447

QY      463  AGSELAKLEMAVFIIHLVLKFWMELEADDP 493
          | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db      448  VGLPLGRHMRHNFALGSLLEKEFWELPDGVP 478

```

```

RESULT 13
US-09-499-302A-2
: Sequence 2, Application US/09499302A
: Patent No. 6369212
: GENERAL INFORMATION:
: APPLICANT: BODUNG-JUN, OH
: APPLICANT: MOON, KYUNG KO
: TITLE OF INVENTION: A CYTOCHROME P450 GENE HIGHLY EXPRESSED IN THE
: TITLE OF INVENTION: INCOMPATIBLE INTERACTION
: FILE REFERENCE: 10324/P64443USO
: CURRENT APPLICATION NUMBER: US/09/499,302A
: CURRENT FILING DATE: 2000-02-07
: NUMBER OF SEQ ID NOS: 10
: SOFTWARE: patentlin Ver. 2.1
: SEQ ID NO 2
: LENGTH: 502
: TYPE: PRY
: ORGANISM: Capsicum annuum
: US-09-499-302A-2

```

[illegible]

```

US-09-126-420A-17
; Sequence 17, Application US/09126420A
; Patent No. 6376753
; GENERAL INFORMATION:
; APPLICANT: BATARD, YANNICK
; APPLICANT: ROBINEAU, TIBURCE
; APPLICANT: DURST, FRANCIS
; APPLICANT: WERCK-REICHART, DANIELE
; APPLICANT: DIDIERJEAN, LUC
; TITLE OF INVENTION: PURIFIED CYTOCHROME P450 CYP76B1 FROM HELIANTHUS
; TITLE OF INVENTION: TUBEROSUS AND ITS APPLICATIONS AS BIOCATALYST IN
; TITLE OF INVENTION: PARTICULAR FOR THE DEGRADATION OF ENVIRONMENTAL
; TITLE OF INVENTION: POLLUTANTS AND FOR ALTERING THE RESISTANCE OF PLANTS S
; FILE REFERENCE: 03/15.0032
; CURRENT APPLICATION NUMBER: US/09/126,420A
; CURRENT FILING DATE: 1998-07-30
; PRIOR APPLICATION NUMBER: 60/054,351
; PRIOR FILING DATE: 1997-07-31
; NUMBER OF SEQ ID NOS: 27
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 17
; LENGTH: 467
; TYPE: PRT
; ORGANISM: Solanum melongena
US-09-126-420A-17

```

[illegible]

QY	456	GGSPRLCAGSELAKEMAYFHHVLTKFWMLAEDDOEFARFENDRGSR	506
		: :	
Db	437	GAGRRCPCG-----MKFG--LANYGRLAOLLYHFDRKIP	469
RESULT 14			
		RESULT 15	
		US-09-126-420A-26	
		; Sequence 26, Application US/09126420A	
		; Patent No. 6376753	
		; GENERAL INFORMATION:	
		; APPLICANT: BATARD, YANNICK	

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: June 10, 2003, 21:57:53 ; Search time 32 Seconds

(without alignments)
1655.071 Million cell updates/sec

Title: US-09-502-426b-2

Sequence: 1 MFETHTLPLLLPSLLS.....FAFPVDFPGLPIRVSRL 513

Scoring table:

BLOSUM62
GapOP 10.0 , Gapext 0.5

Searched: 392085 seqs, 103240269 residues

Total number of hits satisfying chosen parameters: 392085

Minimum DB seq length: 0
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%

Maximum Match 100%
Listing first 45 summaries

Database :

1: /cgn2_6/ptodata/2/pubpaa/US08_NEW_PUB pep: *
2: /cgn2_6/ptodata/2/pubpaa/PCT_NEW_PUB pep: *
3: /cgn2_6/ptodata/2/pubpaa/US06_NEW_PUB pep: *
4: /cgn2_6/ptodata/2/pubpaa/US07_NEW_PUB pep: *
5: /cgn2_6/ptodata/2/pubpaa/US07_PUBCOMB pep: *
6: /cgn2_6/ptodata/2/pubpaa/US07_PUBCOMB pep: *
7: /cgn2_6/ptodata/2/pubpaa/PCRMUS_PUBCOMB pep: *
8: /cgn2_6/ptodata/2/pubpaa/US09_PUBCOMB pep: *
9: /cgn2_6/ptodata/2/pubpaa/US09_PUB pep: *
10: /cgn2_6/ptodata/2/pubpaa/US10_PUBCOMB pep: *
11: /cgn2_6/ptodata/2/pubpaa/US10_PUB pep: *
12: /cgn2_6/ptodata/2/pubpaa/US10_PUBCOMB pep: *
13: /cgn2_6/ptodata/2/pubpaa/US60_NEW_PUB pep: *
14: /cgn2_6/ptodata/2/pubpaa/US60_PUBCOMB pep: *

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Match	Query Length	DB ID	Description
1	816.5	30.5	486	9	US-10-142-231-92 Sequence 92, Appl
2	597.5	22.3	485	9	US-10-142-231-88 Sequence 88, Appl
3	596.5	22.2	497	9	US-10-142-231-64 Sequence 64, Appl
4	592	22.1	496	9	US-10-142-231-90 Sequence 90, Appl
5	588	21.9	509	9	US-10-142-231-67 Sequence 67, Appl
6	585	21.8	507	9	US-10-142-231-65 Sequence 65, Appl
7	583	21.7	484	9	US-10-142-231-57 Sequence 57, Appl
8	571	21.3	501	9	US-10-142-231-58 Sequence 58, Appl
9	569.5	21.2	493	9	US-10-142-231-89 Sequence 59, Appl
10	568.5	21.0	500	9	US-10-142-231-56 Sequence 56, Appl
11	564	21.0	484	9	US-10-142-231-61 Sequence 61, Appl
12	558.5	20.8	512	9	US-10-142-231-62 Sequence 62, Appl
13	558.5	20.8	498	9	US-10-142-231-87 Sequence 87, Appl
14	553.5	20.6	503	9	US-10-142-231-63 Sequence 91, Appl
15	534.5	19.9	492	9	US-10-142-231-60 Sequence 63, Appl
16	458	17.1	491	9	US-10-210-965-4 Sequence 40, Appl
17	302.5	11.3	516	9	US-10-067-534-3 Sequence 3, Appl
18	292	10.9			

20	290.5	10.8	514	10	US-09-947-027-4	Sequence 4, Appl
21	290.5	10.8	514	12	US-10-091-009-4	Sequence 4, Appl
22	289	10.6	511	10	US-09-796-256a-4	Sequence 4, Appl
23	285.5	10.6	502	9	US-10-054-988-69	Sequence 69, Appl
24	285.5	10.6	502	10	US-09-739-254-69	Sequence 69, Appl
25	285.5	10.6	502	10	US-09-904-615-69	Sequence 69, Appl
26	275	10.3	496	9	US-10-067-668-10	Sequence 10, Appl
27	275	10.3	496	9	US-10-175-696-10	Sequence 10, Appl
28	275	10.3	496	9	US-09-945-301-4	Sequence 4, Appl
29	273	10.2	503	10	US-10-146-575-2	Sequence 2, Appl
30	272	10.1	470	9	US-10-067-668-12	Sequence 12, Appl
31	272	10.1	470	9	US-10-175-696-12	Sequence 12, Appl
32	264	9.8	512	9	US-10-142-231-66	Sequence 66, Appl
33	262.5	9.8	514	9	US-10-201-213-5	Sequence 2, Appl
34	258.5	9.6	574	9	US-10-236-433-2	Sequence 2, Appl
35	258	9.6	514	9	US-10-142-231-68	Sequence 68, Appl
36	256.5	9.6	520	10	US-09-992-901-2	Sequence 2, Appl
37	251	9.4	515	10	US-09-796-138-19	Sequence 19, Appl
38	251	9.4	515	10	US-09-909-903-19	Sequence 19, Appl
39	251	9.4	544	9	US-10-067-668-8	Sequence 8, Appl
40	251	9.4	544	9	US-10-175-696-8	Sequence 8, Appl
41	249	9.3	503	10	US-09-796-138-18	Sequence 18, Appl
42	249	9.3	503	10	US-09-909-903-18	Sequence 18, Appl
43	249	9.3	503	10	US-09-957-997-3	Sequence 3, Appl
44	249	9.3	509	10	US-09-817-184-4	Sequence 4, Appl
45	247.5	9.2	493	9	US-10-103-520-2	Sequence 2, Appl

ALIGNMENTS

RESULT 1	US-10-142-231-92	Sequence 92, Application US/10142231
Publication No.	US2003007796A1	
GENERAL INFORMATION:		
APPLICANT:	Croteau, Rodney et al.	
TITLE OF INVENTION:	CYTCHROME P450 OXYGENASES AND THEIR USES	
FILE REFERENCE:	62773	
CURRENT APPLICATION NUMBER:	US/10/142,231	
CURRENT FILING DATE:	2002-05-08	
PRIOR APPLICATION NUMBER:	60/165,250	
PRIOR FILING DATE:	1999-11-12	
NUMBER OF SEQ ID NOS:	95	
SOFTWARE:	PatentIn Ver. 2.1	
SEQ ID NO 92		
LENGTH:	486	
TYPE:	PRT	
ORGANISM:	Taxus cuspidata	
US-10-142-231-92		
Query Match	30.5%; Score 816.5; DB 9; Length 486;	
Best Local Similarity	35.1%; Pred. No. 2.2e+58;	
Matches	176; Conservative 97; Mismatches 184; Indels 45; Gaps 8;	
QY	19	LSLLFLTLTKRRNRK-----TFNLPKSGSGWPGFETIGYKPYATATLSPMOOHV 72
DB	20	LFICVILLRLRRSDROGNSANKPKLPFGSAGLPFGITRTLRDAKSPERRKFFDEHE 79
QY	73	SKYKTIYSNLFGER-TIVSDAGLNFRITLIONBGRLEFCYSYPSISGIGLKMMLVLVD 131
DB	80	LRYGPIFCSTLGRTRAVSVDPFNKYVLONBGRLEFSNALAFRNLIKGYGSAVOGE 139
QY	132	MHRDRSISLNFSLHARLRTLLKDVSEHTLFVLDSONOOSIFSAODEAKKTFTNLMAXH 191
DB	140	LQRHLHTAVVLLKHLFLSSDFMEDIDITIQAGKRWEEGGDIPIOKCNQVINTLMAKR 199
QY	192	IMSNDPEEETEQKKEYVTFMKGVSAPINLPDTAYAKALOSRATTLKFERKMERKL 251
DB	200	LLDLP-SEMGHGYKAFDFGVGAVLSFPIINQGTYARGIRAGILKIKHCICERR- 257
QY	252	DIKEEDDEEVEVTEDEAKMSKDHVAKKOTDDDLGAVLKHNLTSEQILDLITSLPLA 311

Db 258 -----BHPV-----LRNDLTILVREGTSDIELADTILFFVFA 292

Qy 312 GHETSSVAIALAIFLQACPAVEELREHELEIARAKKEGSELNMDYKMDPTQCVI 371

Db 293 GVEISAMATFAVAVKLANENPRALEHDAIILAKAKG-GNEKLTMDYQSMKRVHCYI 351

Qy 372 NETRLGAVNVPFLHKKALKDVRKYDIPSGKVLPLVSAVHLDSRDPRLPMPRMQ 431

Db 352 NETRLGATVVLREAKQDVKVDFIPKGTVSFVSAFHVDAKRYHEADKFLPMRMQ 411

Qy 432 QONNGASSGSGSFSTWGNNTMPFGCGPRLCAGSELALLENMAVFIHLVLFKPMLEADD 491

Db 412 NE-----GQETLEPCYMPFGRGRLCPGLALFELALFLHNVYTKRMQLEID 462

Qy 492 QPFAFPFVDPNGPLPIRY-SRI 512

Db 463 RAYFPPLPSTENGPIRLYSRV 484

RESULT 2

US-10-142-231-88

; Sequence 88, Application US/10142231

; Publication No. US2003007796A1

; GENERAL INFORMATION:

; APPLICANT: Croteau, Rodney et al.

; TITLE OF INVENTION: CYTOCHROME P450 OXYGENASES AND THEIR USES

; FILE REFERENCE: 62773

; CURRENT APPLICATION NUMBER: US/10/142,231

; CURRENT FILING DATE: 2002-05-08

; PRIOR APPLICATION NUMBER: 60/165,250

; PRIOR FILING DATE: 1999-11-12

; NUMBER OF SEQ ID NOS: 95

; SOFTWARE: PatentIn Ver. 2.1

; SEQ ID NO 88

; LENGTH: 485

; TYPE: PRF

; ORGANISM: Taxus cuspidata

US-10-142-231-88

Query Match 22.3%; Score 597.5; DB 9; Length 485;

Best Local Similarity 27.9%; Pred. No. 1.6e-40;

Matches 147; Conservative 114; Mismatches 191; Indels 75; Gaps 13;

Qy 9 LRLPLLSL-----LSLLFLFLILKRRNRKTRFNLPPKSGMPLFGTIGLKPYATTATL 64

Db 4 LKOLEVSPSLFVTLVAVAGIILFFRSKRHSVXKLPNGNLGPPVGLGTLQVLRSGSSTP 63

Qy 65 GDFMOQVSKYKTYRSLNFGPTVSADAGLNRFILQNGRLFCFSYPRSIGIIGKWS 124

Db 64 QQFLERMSKRGDVKFSIIGHPTVVLGCPAGNRLVLSNENKLVOMSPSMKLLIGEDC 123

Qy 125 MLVIVGDMHRDMSISLFLSHARLRTILK--DVERHTLFLVDSMQ-----Q 170

Db 124 LGKRTGQHRIYRALTRFLGRLQALQNHFAKMSGIOH--INEMKMGKDEAVPLVYK 180

Qy 171 NSISADDEAKKFTFNLMKAKIMSMDPEEETEOUKKEYVTFMKGVSAPLNLPGTAYHK 230

Db 181 DLVFSV--ASRLFFGTIEHLQ-----EQLHMLEVILVGSFSPINIGFGSYHK 228

Qy 231 ALQSRATILKFIERKMEERKLDIKEDEEVEVTEDEAEKSKSDHVRKQRTDDLLGMV 290

Db 229 ALQKRALDADMTILKRRRELKAGTASENO-----DLSVL 266

Qy 291 L-----KHSNLSTEQIIDLILSLFAGHETSSVAIALAIFLQACPAVEELREHELEIA 345

Db 267 LTFDERGNSLADKEILDNFSMLHGSYDSTNSPLMLIKVLASHPSYEVKVADEQGITL 326

Qy 346 RAKKELESELNMDYKMDPTQCVINETLRLGNVVRFLHKKALKDVRKYGDIPIPSGKV 405

Db 327 STKME-GE-EIAMDRLKEMKYSQVQVETLRMPPIGTFGRKATIDIHNYGTYTIPKQKL 384

Qy 406 LPLVSAVHLDSRDPRLPMPRMQOONNGASSGSGSFSTWGNNTMPFGCGPRLCAGS 465

Db 385 LMTTSTQTKREYFKADQPKRSRFEDEGKHVP-----YTLPLFGGGMVPCGW 434

Qy 466 ELAKLEMAVFIHLVLFKPMLEAD--DOPFAFPFVDP-NGLPPIRY 509

Db 435 EFARKMETLFLHHPVAFSGKLALIDPENLISGRPLRPLVNGPLPIKL 481

RESULT 3

US-10-142-231-64

; Sequence 64, Application US/10142231

; Publication No. US2003007796A1

; GENERAL INFORMATION:

; APPLICANT: Croteau, Rodney et al.

; TITLE OF INVENTION: CYTOCHROME P450 OXYGENASES AND THEIR USES

; FILE REFERENCE: 62773

; CURRENT APPLICATION NUMBER: US/10/142,231

; CURRENT FILING DATE: 2002-05-08

; PRIOR APPLICATION NUMBER: 60/165,250

; PRIOR FILING DATE: 1999-11-12

; NUMBER OF SEQ ID NOS: 95

; SOFTWARE: PatentIn Ver. 2.1

; SEQ ID NO 64

; LENGTH: 497

; TYPE: PRF

; ORGANISM: Taxus cuspidata

US-10-142-231-64

Query Match 22.2%; Score 596.5; DB 9; Length 497;

Best Local Similarity 29.4%; Pred. No. 2e-40;

Matches 152; Conservative 98; Mismatches 200; Indels 67; Gaps 12;

Qy 16 PSLSLLL-----FLILKRRNRKTRFNLPPKSGMPLFGTIGLKPYATTATLQDFMQ 69

Db 21 PALSLTLAPLAIILILLFLRYNHRSSVXKLPNGNLGPPVGLGTLQVLRSGSSTP 80

Qy 70 QHVSRYGKIYRSLNFGPTVSADAGLNRFILQNGRLFCFSYPRSIGIIGKWSMLVLY 129

Db 81 DRLKRFGPVMTSLGHPTVVLGCPAGNRLVLSNENKLVOMSPSMKLLIGEDSIYARR 140

Qy 130 GDMHRDMSISLFLSHARLRTIL--LKDVERHTLFLVDSMQOONISFSAODEAK----- 181

Db 141 GEDHILKRALARFLGAGALQVLYLGRMSEIGHF-----NEKMGKDEAVKPLVY 191

Qy 182 -KFTNLMKAKIMSMDPEEETEOUKKEYVTFMKGVSAPLNLPGTAYHKALQSRATYIK 240

Db 192 RLILSISLSTLFVVDNDGHQ-Q-KQLHMLETILVGSILVPLDPGTRGRKGLQARLKDE 250

Qy 241 FIERKMEERKLDIKEDEEVEVTEDEAEKSKSDHVRKQRTDDLLGMV-----KHSN 295

Db 251 ILSLILKRRRDLKSGIASD-----DQDLSVLTLFPRDEKGS 288

Qy 296 LSTEQIIDLILSLFAGHETSSVAIALAIFLQACPAVEELREHELEIARAKKEGSE 355

Db 289 LITDQIILNFSAMFASVDTYAPRALILFKLLXSNPEYENKFOQDLEILGKKE-GE-E 346

Qy 356 LNMDDYKMDPTQCVINETLRLGNVVRFLHKKALKDVRKYGDIPIPSGKVPLVSAVHL 415

Db 347 ISMKDLKSMKYTWQVQVQVETLRMPPIGTFGRKATIDIHNYGTYTIPKGMVLCSPYTTTH 406

Qy 416 NSRYQVPLVFNPMRMQOONNGASSGSGSFSTWGNNTMPFGCGPRLCAGSELAKLEMAV 475

Db 407 EEPYEPPEPFRSREDGGRHVP-----YTVVFGGGLRTCPGMEFSKIEILLF 456

Qy 476 IHLVLFKPMLEAD--DOPFAFPFVDP-NGLPPIRY 509

Db 457 VHHFVKNSSYIIPVDPNKKVLSLDPPLPANGFSIKL 493

RESULT 4

US-10-142-231-90

; Sequence 90, Application US/10142231

; Publication No. US2003007796A1

; GENERAL INFORMATION:

Page 3

Best Local Similarity 30.6%; Pred. No. 1.8e-39;
Matches 158; Conservative 94; Mismatches 204; Indels 60; Gaps 14;

```

Db      14 LPFLTP-LIGALVLLHFRHRNRS--VKLPKGLGPEVIGETIQFIRALRSQPPQKFD 71
QY      70 QHVSXKGIYRSNLFGEPTIVSADAGLNRFILQNEGRLEFECYSRISIGILGKSMVLV 129
Db      72 DRVQKFGGVRTSLIGNPVLVWCGPACNRVLVSNEDKLYQLEARNLSLMLKMGOSLAKR 131
QY      130 GDMHRMRSISLNFSLHARLTLLKDOVERHTLVLDWQONSIFSAODEAKKTFENIMA 189
Db      132 QEDHRTLRALARLGLFOALXNYMTKISSRTEHHMNEKWK-----GKDEVYR--TLPLIR 183
QY      190 KHMMSMDPG-----EETEQLKKEVYTPMKGVVAPLNPGTAHKAQSRATILKFI 242
Db      184 ELIFSNASSLPEDINDHQERLHLLLEAVVGSMSIPDPGTRLRALQARSKLDEL 243
QY      243 ERKMEERKLDIKEEDQEEVEKTEDEAEMSKSDHVRKQRTDDLLGWL-----KHSNLS 297
Db      244 SSLKSRKRL-----VSGIASD-----DQDLISVLLTKRDERGNFLT 281
QY      298 TEQILDLLILSLFAGHETSSVALAIFLQACPKAVEELREHELEIARAKKEGESELN 357
Db      282 DKEILDNFSLLHASYDTVSPVLTLLKLSNDECEKVVQEQGIY-ANKRIGE-EIS 339
QY      358 WDOYKKMDPQCYNETLRGLGNVRFLLHRKALKDVRKGYDIPSGMKVLPVISAHLDNS 417
Db      340 WKDLKAKVYTWQVQETLRMPPLGSGFRKAMVDIDGYTIPKGMMLTWTGTHLEE 399
QY      418 RYDOPNLFNPMWQOONNGASSSGSFSTGNVMPFGPRLCAGSELAKLEMAVFIH 477
Db      400 YFNPLKFRPSRFE-----DGRVTPY--TFIFPGGARGCPGMEFSKTEILLFIH 448
QY      478 HLVKFNWELAE--DQPAFPEVDFP-NGLPRIVS 510
Db      449 HFVTFSSYLPVDSNEKISADPEFPPLPANGFSIKLS 484

RESULT 7
US-10-142-231-57
; Sequence 57, Application US/10142231
; Publication No. US2003007796A1
; GENERAL INFORMATION:
; APPLICANT: Croteau, Rodney et al.
; TITLE OF INVENTION: CYTOCHROME P450 OXYGENASES AND THEIR USES
; FILE REFERENCE: 62773
; CURRENT APPLICATION NUMBER: US/10/142,231
; CURRENT FILING DATE: 2002-05-08
; PRIOR APPLICATION NUMBER: 60/165,250
; PRIOR FILING DATE: 1999-11-12
; NUMBER OF SEQ ID NOS: 95
; SOFTWARE: Patentln Ver. 2.1
; SEQ ID NO 57
; LENGTH: 484
; TYPE: PRT
; ORGANISM: Taxus cuspidata
US-10-142-231-57

```

```

Query Match      21.7%; Score 583; DB 9; Length 484;
Best Local Similarity 29.3%; Pred. No. 2,4e-39;
Matches 154; Conservative 94; Mismatches 209; Indels 68; Gaps 13;

```

```

QY      5 EHTTLPLDLLLPLSLTLLFLILL--KRRNRKTRFNLPRKSGSPFLGETIGYILKPRTA 61
Db      4 EMOTFVQLESSPVLSTLTLTLILFLIFCSKQYRSKSLRPGNMGFPLIGETIA---LAS 59
QY      62 TTAGDFMOQHVSKYKLYRSNLFGEPTIVSADAGLNRFILQNEGRLEFECYSRISIGILG 121
Db      60 QTDKPFQDMKKKFGKFKTSILGHPTIVLCGSSGNFLLSNEKVLVAMPFPNSSSKILG 119
QY      122 KMSMLVGVDMHRMRSISLNFSLHARLTLLK---DVERHTLVLDWQONSIFSAOD 178
Db      120 QDSVLGIGBEHRIVPRALARCGLQALQNVVSKMSSEIORH---INQKMGKGEVVKLP 176
QY      179 EAKFTYTNLAKHMSMDPGEETEQLKKEYVTFMKGVVAPLNPGTAHKAQSRATI 238

```

```

Db      177 LIRSLVFSIANSLEFGI--TDEQOQERLHLLLETVTGLCIPLDPEGTFRRALHARSKL 235
QY      239 LKFIKRMEERKLDIKEEDQEEVEKTEDEAEMSKSDHVRKQRTDDLLGWL-----KH 293
Db      236 DEIMSSVIERRRNDR-----LGAASSODDLISVLLTFKREDEG 273
QY      294 SNLSTEQILDLLILSLFAGHETSSVALAIFLQACPKAVEELREHELEIARAKKEGE 353
Db      274 NPFAKTEILDNFSLLHAYDTTISPLTVLFKLVSSNDECEYENIAOEOLEILRNKKD-GE 332
QY      354 SELMNDVKKMDPQCYNETLRGLGNVRFLLHRKALKDVRKGYDIPSGMKVLPVISAHV 413
Db      333 -DISWADLKMKVYTWQVQETLRMCPRYAGNFRKALTIDHGYTIPKGMWILCSPTYTH 391
QY      414 LDNSYDOPNLFNPMWQOONNGASSSGSFSTGNVMPFGPRLCAGSELAKLEMA 473
Db      392 SKEEFIDPEKFRPSRFEQGRDVA-----YFIFPGGGLRCPGMEFAKMEIL 441
QY      474 VFIIHLVLKF-----NWELEDDQPAFPEVDFP-NGLPRIV 509
Db      442 VFMHFVKAFFSFPVDPNEKISTD-----PLPSIPVNGFSINL 480

```

```

RESULT 8
US-10-142-231-58
; Sequence 58, Application US/10142231
; Publication No. US2003007796A1
; GENERAL INFORMATION:
; APPLICANT: Croteau, Rodney et al.
; TITLE OF INVENTION: CYTOCHROME P450 OXYGENASES AND THEIR USES
; FILE REFERENCE: 62773
; CURRENT APPLICATION NUMBER: US/10/142,231
; CURRENT FILING DATE: 2002-05-08
; PRIOR APPLICATION NUMBER: 60/165,250
; PRIOR FILING DATE: 1999-11-12
; NUMBER OF SEQ ID NOS: 95
; SOFTWARE: Patentln Ver. 2.1
; SEQ ID NO 58
; LENGTH: 501
; TYPE: PRT
; ORGANISM: Taxus cuspidata
US-10-142-231-58

```

```

Query Match      21.3%; Score 571; DB 9; Length 501;
Best Local Similarity 29.2%; Pred. No. 2,4e-38;
Matches 150; Conservative 104; Mismatches 200; Indels 60; Gaps 14;

```

```

QY      12 LLLPLSLLLFLILKRRNRKTRFNLPRKSGSPFLGETIGYILKPTATTLGDPMOQH 71
Db      28 LLSLAFTAVALIFVLLFRKSRPSTNPPGNGFPFGETIGYILRALRSESPHMFDFDR 87
QY      72 VSKYKGIYRSNLFGEPTIVSADAGLNRFILQNEGRLEFECYSRISIGILGKSMVLVGD 131
Db      88 LKRGKRVKTSITGHPTAVFCGPACNRRTIYSEHLLVSSGNSVVKLVGOSIYTKTGE 147
QY      132 MHRMRSISLNFSLHARLTLLKDOVERHTLVLDWQONSIFSAODEAKKFT 184
Db      148 EHRIFLGLVNEFLGPHALQSTYPMKSKIOENIKH-----WKGXDEVNMLPSIRQV 200
QY      185 FNLAKHMSMDPGEETEQLKKEYVTFMKGVVAPLNPGTAHKAQSRATILKFI 244
Db      201 FSISSSLEFDIN-EDQOQERLHLLLETVTGLCIPLDPEGTFRRALRARSKLDELISR 259
QY      245 KMEERKLDIKEEDQEEVEKTEDEAEMSKSDHVRKQRTDDLLGWL-----KHSNLS 299
Db      260 LIESRRKDMRGSIAS-----ISKNLISVLLAFKDERGNPLDVT 297
QY      300 QILDLLILSLFAGHETSSVALAIF-FLQACPKAVEELREHELEIARAKKEGESELN 358
Db      298 EILDNFSMLHASYDT-TVSPVLCIFKLISANPECEKVVQEQGIYILNKKD-GE-EKCM 354
QY      359 DDYKMDPQCYNETLRGLGNVRFLLHRKALKDVRKGYDIPSGMKVLPVISAHLNSR 418

```


Db 355 NDKAKKTYWQAAQETWRLFPFAGSGFRKYADIIHGYIIPGKAMATNYSRKEY 414
 QY 419 YDOPNLEPWRMOOONNGASSGSESTWNNYMPGGGPRCAGSELAKEMAVRTH 478
 Db 415 FDEPDMFKPSRF-----GDCKYVA-PYTFPLPGAGIRICPGMEFAKLEMLFIH 463
 QY 479 LVKEMWEIAED--DOPFAFPVDFP-NGLPPIRV 509
 Db 464 FVKNFSGYLPIDTKERISGDPFPLPKNGFPIKL 497

RESULT 9

US-10-142-231-89
 ; Sequence 89, Application US/10142231
 ; Publication No. US2003007796A1
 ; GENERAL INFORMATION:
 ; APPLICANT: Croteau, Rodney et al.
 ; TITLE OF INVENTION: CYTOCHROME P450 OXYGENASES AND THEIR USES
 ; FILE REFERENCE: 62773
 ; CURRENT APPLICATION NUMBER: US/10/142,231
 ; CURRENT FILING DATE: 2002-05-08
 ; PRIOR APPLICATION NUMBER: 60/165,250
 ; PRIOR FILING DATE: 1999-11-12
 ; NUMBER OF SEQ ID NOS: 95
 ; SOFTWARE: PatentIn Ver. 2.1
 ; SEQ ID NO 89
 ; TYPE: PRF
 ; ORGANISM: Taxus cuspidata
 US-10-142-231-89

Query Match 21.2%; Score 569.5; DB 9; Length 493;
 Best Local Similarity 28.9%; Pred. No. 3.1e-38;
 Matches 146; Conservative 101; Mismatches 207; Indels 51; Gaps 11;

QY 18 LLSLLFLILL--KRRNRKTRFNLPPGSGWPLGETIGYLAQPYATTGDMQOHVSKY 75
 Db 23 ILSLTALILEFFFRYRNKSSHKLPGNIGFPFGETIOFLRSLSOTPEFFDERVKKF 82
 QY 76 GKIRSNLFGPEPTIVADAGLNRFILONEGRLEFCSPRSIGILGKMSMLVVGDMHRD 135
 Db 83 GPVFKTSILGAPFVIVICGAGSKLVLSNEDKLVOMESPSSLKLMGNSLILYREEHRT 142
 QY 136 MNSISLNLISHALRTILK--DVERHTLVLDSSQONSIFSAODEAKFTFNMAKHI 192
 Db 143 LRSALSRFLGPOLQTYIAKMSIEIRH--INEKKKGREYVTLPLIRGLVSIASSLF 199
 QY 193 MSMDPEETEOUKRYVFMKGVSAPLNPGTAVHKALQSRATILKFIERKMEERKLD 252
 Db 200 FDIIN-DEPOERLHHLHESLVAGSMAVRLDPEPTRKRAVEARSKIDEALHSLIKRSRD 258
 QY 253 IKEEDDEEVEVTEDEAEMSKSDHVAKORTDDLLGMVLKHSN-----LSTEQILLILS 307
 Db 259 LLS-----GKASSNODLSVLLSEKDERGNPLADEEILLNFSL 296
 QY 308 LFPAGHETSSVAIALAIFLOACPKAVEELREEHLEIARAKKELGSEELNMDYKKMDF 367
 Db 297 ILHASVDTITSPVVLTKLLSNPECYDYVQOFOELIANKKR-GE-EISMKRLKMKYI 354
 QY 368 QCVINETLRLGNVVRFLHKKALKDVRKGYDIPSGMKVLPVISAHLNDSRYDOPNLEP 427
 Db 355 WGVVDETLLMFPLPFLFSGFRKAMVDINYGDTIPKGIIVLMTYSTHVKEEYFNEPKFRF 414
 QY 428 WRMOOONNGASSGSESTWGNVMPFGGPRLCAGSELAKIEMAVFIHHLVLRKNWEL 487
 Db 415 SREHIDGRVAP-----YTFLPFGGGLTCGMEFSEKTEILLFIHFVKTSYSL 464
 QY 488 AED--DOPFAFPVDFP-NGLPPIRV 509
 Db 465 PVDPMKISADPEPPPLPANGFSIKL 489

RESULT 10

US-10-142-231-59
 ; Sequence 59, Application US/10142231
 ; Publication No. US2003007796A1
 ; GENERAL INFORMATION:
 ; APPLICANT: Croteau, Rodney et al.
 ; TITLE OF INVENTION: CYTOCHROME P450 OXYGENASES AND THEIR USES
 ; FILE REFERENCE: 62773
 ; CURRENT APPLICATION NUMBER: US/10/142,231
 ; CURRENT FILING DATE: 2002-05-08
 ; PRIOR APPLICATION NUMBER: 60/165,250
 ; PRIOR FILING DATE: 1999-11-12
 ; NUMBER OF SEQ ID NOS: 95
 ; SOFTWARE: PatentIn Ver. 2.1
 ; SEQ ID NO 59
 ; TYPE: PRF
 ; ORGANISM: Taxus cuspidata
 US-10-142-231-59

Query Match 21.2%; Score 568.5; DB 9; Length 500;
 Best Local Similarity 28.9%; Pred. No. 3.9e-38;
 Matches 150; Conservative 103; Mismatches 195; Indels 71; Gaps 14;

QY 16 PSLISLIL-----FILLKRRNRKTRFNLPPGSGWPLGETIGYLAQPYATTGDMQ 69
 Db 24 PAILSTALTAAGIYLVITSKRSSLAKLPGLGIFGETIEFYKALRSPTLRQFVE 83
 QY 70 QHVSYKGIKIRSNLFGPEPTIVADAGLNRFILONEGRLEFCSPRSIGILGKMSMLV 129
 Db 84 EREGKRGVFKTSILGKPFVILCGPAONRLVLSNEEKILHVSQAQIARILGINSVAVVR 143
 QY 130 GDMHRDNRSSIMPLSHARLRTILK--VERHTLVLDSSQ--ONSIFS-----AOD 178
 Db 144 GDHRLVRLVLAGFLGAGLQYIGKMSALIRN--INRKKGKDVNVLSTVRLVMD 200
 QY 179 EAKKFTFNMAKHIMSMDPEETEOUKRYVFMKGVSAPLNPGTAVHKALQSRATI 238
 Db 201 NSAILFENIVDK-----ERKQOLHEIKITIIASHFGIPLINPEFLRKALGSLKR 251
 QY 239 LKFIERKMEERKLDIKEEDDEEVEVTEDEAEMSKSDHVAKORTDDLLGMVLKHSN-- 295
 Db 252 KKLISLALKERKDELRS-----KLIASSNODLSVLLSEFRBERG 289
 QY 296 --LSTEQILLILSIFAGHETSSVAIALAIFLOACPKAVEELREEHLEIARAKKELG 353
 Db 290 KPLSDEAVLDNCAMLDASTDITTSQMTLILKLSNPECFEKVVOOLEIASNKE-GE 348
 QY 354 SELNMDYKKMDFTOCVINETLRLGNVVRFLHKKALKDVRKGYDIPSGMKVLPVISAH 413
 Db 349 -EITMDIKAMKKTWQVLQSLMLSPVFTLRTKTMNDINHGYTIPKGMVWYWTYSTH 407
 QY 414 LNSRVDOPNLEPWRMOOONNGASSGSESTWGNVMPFGGPRLCAGSELAKIEMA 473
 Db 408 QKDIYFQOPKFMPSREED-----GHLDAY--TFVPEFGGGRTCGMEYAKVEIL 457
 QY 474 VFIHHLVLRKNWELAED--DOPFAFPVDFP-NGLPPIRV 509
 Db 458 LFLHFFVKAFSGYTPDPHERICGYVPLVVKGFPIKL 496

RESULT 11

US-10-142-231-56
 ; Sequence 56, Application US/10142231
 ; Publication No. US2003007796A1
 ; GENERAL INFORMATION:
 ; APPLICANT: Croteau, Rodney et al.
 ; TITLE OF INVENTION: CYTOCHROME P450 OXYGENASES AND THEIR USES
 ; FILE REFERENCE: 62773
 ; CURRENT APPLICATION NUMBER: US/10/142,231
 ; CURRENT FILING DATE: 2002-05-08
 ; PRIOR APPLICATION NUMBER: 60/165,250
 ; PRIOR FILING DATE: 1999-11-12

```

; NUMBER OF SEQ ID NOS: 95
; SOFTWARE: PatentIn Ver. 2.1.1
; SEQ ID NO 56
; LENGTH: 484
; TYPE: PRT
; ORGANISM: Taxus cuspidata
US-10-142-231-56

```

Query Match	21.0%;	Score 564;	DB 9;	Length 484;	.
Best Local Similarity	28.5%;	Pred. No. 8.6e-38;			
Matches 146;	Conservative 95;	Mismatches 197;	Indels 74;	Gaps 12	

```

0Y      HTLLPDL---LPSSLSTLLPFLTLKRRNRKTFEFPNPGSGMPLCEGTIGYKLPYAT 63
1       | 1111 |---| 1111 |---| 1111 |---| 1111 |---| 1111 |---| 1111 |
Db      7 HESSPLLLSTLAVILGTLTLLLSKQYRRSR-KLRPMGMGFLIGETIALI----SDT 61

0Y      64 LDFPMQOAVSKGKGYRSLFGEPTVSDAGLNRTILONEGRHLEFCSYRSTIGIGLKW 123
1       | 1111 |---| 1111 |---| 1111 |---| 1111 |---| 1111 |---| 1111 |
Db      62 PRKFIDDRKKGKGLVFKTSLSHPAVVICGSSANRPLLSNEEKLVRNLSPEANAULKLGOD 121

0Y      124 SMLVYGDHRRMRSISLNFSLHARLTILK---DYERTRLPVLDSMOQNSIFYAODEA 180
1       | 1111 |---| 1111 |---| 1111 |---| 1111 |---| 1111 |---| 1111 |
Db      122 CYMGKTYVEHGIYRVALARALORPALQNVAAKSSIEIHHI-----NQMKKGDEY 177

0Y      181 K-----KFTENLMAKHNSMDPGEETEQLKKEYVTVMKGVSADPLNLPYAHKALOS 234
1       | 1111 |---| 1111 |---| 1111 |---| 1111 |---| 1111 |---| 1111 |
Db      173 KYLPRLRSLVFESISLFEFGIN-DEHQOQRKLNHLLFTVMAGLVSIPLDFEGTFRKALYA 231

0Y      235 RATILKFTERKKEERKLDIKEEDQEEBEVKTEDAEAMKSDHVRKQOTDDLDLGMVY-- 291
1       | 1111 |---| 1111 |---| 1111 |---| 1111 |---| 1111 |---| 1111 |
Db      232 RSKLDEIMSSVYERRRSDRS-----GAASDDDLISLVYTFK 265

0Y      292 --KHSNLSQTLIDLLLSLLFAGHETSVAIALAIFLQACPAVEELREHLEIARAKK 349
1       | 1111 |---| 1111 |---| 1111 |---| 1111 |---| 1111 |---| 1111 |
Db      270 DERGSFADKEILLDNFSEFLHALYDTTISPLTFLFKLSSPCEYENIAQEOELIGNKK 329

0Y      350 ELGSELMNMDVKKKDNDFQTCVNLNETRLGNVAFELRKALKDVRYGYDPLPSMKVLPVI 409
1       | 1111 |---| 1111 |---| 1111 |---| 1111 |---| 1111 |---| 1111 |
Db      330 D--REIISKDKLDMKTYTQAOVETILRMPPRYGYIRFALTDIDYQYITPKMRILCSP 387

0Y      410 SAVHLDNRSKYDOPNLFPWRMOOQONNGASSSGSGSFSTWGNMNNYMPFGGGRILCASLELAK 465
1       | 1111 |---| 1111 |---| 1111 |---| 1111 |---| 1111 |---| 1111 |
Db      388 HTTSHKEEFYDEDEEPRFRSREFDQGRHVP-----YTLIPGGGLRICAGWGFAPK 437

0Y      470 LEMAVFIHHLYLKF-----NWELEADDOP 493
1       | 1111 |---| 1111 |---| 1111 |---| 1111 |---| 1111 |---| 1111 |
Db      438 MEILFEMHFEVTFESHIPVDPEKTSRDLPL 469

```

```

RESULT 12
US-10-142-231-61
: Sequence 61, Application US/10142231
: Publication No. US2003007796A1
: GENERAL INFORMATION:
: APPLICANT: Croteau, Rodney et al.
: TITLE OF INVENTION: CYTOCHROME P450 OXYGENASES
: FILE REFERENCE: 62773
: CURRENT APPLICATION NUMBER: US/10/142,231
: CURRENT FILING DATE: 2002-05-08
: PRIORITY APPLICATION NUMBER: 60/165,250
: PRIOR FILING DATE: 1999-11-12
: NUMBER OF SEQ ID NOS: 95
: SOFTWARE: Patentlin Ver. 2.1
: SEQ ID NO 61
: LENGTH: 500
: TYPE: PRT
: ORGANISM: Taxus cuspidata
: FEATURE:
: NAME/KEY: misc_feature
: LOCATION: (1)..(500)
: OTHER INFORMATION: "aa" equals any peptide
US-10-142-231-61

```

Query Match	21.0%;	Score 564;	DB 9;	Length 500;
Best Local Similarity	28.3%;	Pred. No. 9e-38;		
Matches 145;	Conservative 102;	Mismatches 213;	Indels 52;	Gaps 11;

QY	5	LPLLLPLPLLLFLLLILKRRNKTFFNPPGSGMPFGEITGLKYATTATLDDM	68
Db	26	VLSLSLITILLGVLLMFLLKNG--SSVTLRPGNLGSPFGEITPLRLARSTPQTF	83
QY	69	QQHVSRYGKIYNSNLGFEPTIYSADAGNLRIPLONEGRHLEPCSPNSIGGLKWSMLVL	128
Db	84	DEWYKFFGVFFTRIVGHPTVVLGCPEDGNRFLLSNEKLVQASLPNSSEKIKGYSLSK	143
QY	129	VGMHMDMSISLNFISHARLTILK--OVERHTLEVLDSMOONSIFSAODEAKKTF	185
Db	144	RGEHRILTAALARFLRPOALLGYAKKSSQIHH---IKKKMGNDENVKVLPLRIRLF	200
QY	186	NLMKHHMSMDGEEETDLKKEYTFMKGVYAPLNPCTAYNKALOSRATILKFTERK	245
Db	201	NIASSLFEGIN-DEHQEOLHNLLEAIYGLSIVLPDFPCTRFKALDARSKIDELISSL	259
QY	246	MEERKLDIIEEDQEEBEVKTEDEAEMSKSDHYKROTDLDLGVL-----KHSNLTSEQ	300
Db	260	MESRRRDLGLGASENQ-----DLSVLLFKDERGNPLTDKE	297
QY	301	ILDLLSLFAGHETSSVAALATIFLOACRAVEELREHLEIARAKELGSESLMWD	360
Db	298	IFDNFSFMLHASYDTTVPSTGLMTLLRSPDCEYKELVQEOGLGVNKK-GE-EISMND	355
QY	361	YKMDTQGVINETLRIGVYRFLRKRLKDVRYKGVDPISGKVLVPLVISAVHLNDSRYD	420
Db	356	LKAAKTKCVQVESMMLRPFVFGSYRKAAXYTIHIDGITIFPGMIWIFMSPTTHCKEYFN	415
QY	421	QPNLEPMRWQOONNGASSGSGSFTSGNNYMPFGGGRPLACASELAKEMAFVHNLV	480
Db	416	EADRFMPRSFEEGKYVAPY-----FLRFGAGLRVCPGMEFAKTEILLVHNFI	464
QY	481	LKFNWELAD--DQFAFPYDFP-NGLPIRV	509
Db	465	TTFSSYIPLDPKDISGDFPPLPINGSML	496

```

RESULT 13
US-10-142-231-62
; Sequence 62, Application US/10142231
; Publication No. US20030077796a1
; GENERAL INFORMATION:
; APPLICANT: Croceau, Rodney et al.
; TITLE OF INVENTION: CYTOCHROME P450 OXYGENASES AND THEIR USES
; FILE REFERENCE: 62773
; CURRENT APPLICATION NUMBER: US/10/142,231
; CURRENT FILING DATE: 2002-05-08
; PRIOR APPLICATION NUMBER: 60/165,250
; PRIOR FILING DATE: 1999-11-12
; NUMBER OF SEQ ID NOS: 95
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 62
; LENGTH: 483
; TYPE: PR1
; ORGANISM: Taxus cuspidata
; US-10-142-231-62

```

Query Match	20.8%	Score 558.5	DB 9	Length 483
Best Local Similarity	29.6%	Pred. No. 2.4e-37		
Matches 152; Conservative	91	Mismatches 222	Indels 49	Gaps 12

```

QY      8 TLTPLLPPLSLSTLLL-----FTLLRRRNRKTRFNPSPGSMPLIGETIYLAKRYTA   61
Db      3 TLIQIQSSDFLSFTLAFGLGVVLLIRMYHRNAKRLKPENGLDPIRGETI-----TFAS  58

QY      62 TTLGDPMQOAVSKTKGIYRSNLFCEPTIVYSADAGLNRFILDNERRLEFCESPRISGISILG 121
Db      59 QPPQKFLNLRGRKKFPDPVEKFTSLIGHPVTVALCGSSGNAPLNSNEKLTVRMSLPMSNYMKLLG 118
```

QY 122 KMSALVGDHMRDRMSISLNFLSHARLTILK---DVERHTFLVDSMOONSIFSAD 178
 Db 119 QDSLLGKTQEHRIYRTALGRFLGPOLQNHVAKKSSDIQH---INOKMGNDEYKVL 175
 QY 179 EAKKFTFNLMAKHIMSMPDGEETBOLKEEYTFPKGVYSAFLNPGTAVHKALOSRTI 238
 Db 176 LIRNLVFSATSLFEGIN-DEHQERLHLLETTYMGAVCIPLAPGSGFRKALQARSEL 234
 QY 239 LKFIKKMEERLIDKEEDOEDEEVKTEDEAMSKSDHVRKQRTDDLLGVNLKHSNST 298
 Db 235 DGLISLMLKIRRSRLR-----SGAASSQDILSVLITFKDRG-----NPLTD 277
 QY 299 EQLILILSLFAGHETSSVALAIFLQACPAVEELREHELEIARAKKEGSELNWD 358
 Db 278 KEILDNFVYLHGLDITLPLTLFLKSSNTECEYVVOLEIL-SHREKEE-ELGW 335
 QY 359 DDYKMDPTOCVINETRLGNVYRFLHKKALKDVRKYDIPSGMKVLPVISAHLNDR 418
 Db 336 KDLKSMKYTWQALQETELMPPVYGNFRKALTDIHYDITLIPKMRVLCSPFTTHSNEY 395
 QY 419 YDQNLNFWRMQOONNGASSSGSFTWGNMYPFGGPRLCAGSELAKEMAVFTIH 478
 Db 396 FNEDEFPSPRFEQGGKVP-----YTFIPGGGLRICGMEFAKTEMLFTHY 445
 QY 479 LVLKFMELAD--DOPFAFPVDP-NGLPYR 509
 Db 446 FVTFSSYVPVDPNEKISADPLASFPVNGFSVKL 479

RESULT 14

US-10-142-231-87
 ; Sequence 87, Application US/10142231
 ; Publication No. US20030077796A1
 ; GENERAL INFORMATION:
 ; APPLICANT: Croteau, Rodney et al.
 ; TITLE OF INVENTION: CYTOCHROME P450 OXYGENASES AND THEIR USES
 ; FILE REFERENCE: 62773
 ; CURRENT APPLICATION NUMBER: US/10/142,231
 ; PRIOR FILING DATE: 2002-05-08
 ; PRIOR APPLICATION NUMBER: 60/165,250
 ; PRIOR FILING DATE: 1999-11-12
 ; NUMBER OF SEQ ID NOS: 95
 ; SOFTWARE: Patentln Ver. 2.1
 ; SEQ ID NO 87
 ; LENGTH: 512
 ; TYPE: PRF
 ; ORGANISM: *Taxus cuspidata*
 US-10-142-231-87

Query Match 20.8%; Score 558.5; DB 9; Length 512;
 Best Local Similarity 27.9%; Pred. No. 2,6e-37;

Matches 143; Conservative 108; Mismatches 210; Indels 51; Gaps 12;

QY 10 LPLILPLSLILSLILFLILKRRNRKTRFNLPGKSGMPFLGETIGLYKPYATTLGDMDQ 69
 Db 27 LSVTIVTAIAVITLTLVLRSPQSCVNLPPGKIGYPRIGETILOLQAFRNNRPOQFD 86
 QY 70 OHYSKYGKTYRSLNREPTIVSADAGLNFILQNGRLFECSYPSISGIGILGKWSMLVLY 129
 Db 87 ENOKRFGSVFKTSLIGDRTVLVCGPSGNRLILSNNEKLVKAEASMPSSIKLIGEDSIAGKN 146
 QY 130 GMDHMDRSISLNFLSHARLTILK---DVERHTFLVDSMOONSIFSADQAKKFTFN 186
 Db 147 GEKHRLIRAAVRYLGPALQNYMAKMSSEIHH---MNEKKKGEOYKVLPLVAVENFS 203
 QY 187 LMAKHIMS-DPGEETBOLKEEYTFPKGVYSAFLNPGTAVHKALOSRTIILFIRK 245
 Db 204 IATSLFPGVNDGER-ERLHDLTALAGVSIPLDPGTNYRKALRLKIDKVLSSL 261
 QY 246 MEERLIDKEEDOEDEEVKTEDEAMSKSDHVRKQRTDDLLG-VYL-----KHSULSTEQ 300
 Db 262 IERRSDLRSGVAGSNE-----DLSVWLITPKDEEGNPLTDKE 299

QY 301 ILDLILSLFAGHETSSVALAIFLQACPAVEELREHELEIARAKKEGSELNWD 360
 Db 300 ILDNSTLHASDYTTLSALTFLTKIMSSSTECYHKNVQOELRLVSKKE-GE-ELSKD 357
 QY 361 YKMDPTOCVINETRLGNVYRFLHKKALKDVRKYDIPSGMKVLPVISAHLNDRYD 420
 Db 358 LKDKKTYWQVQETLRMFPLFGSFRKALTDIHYDITLIPKMRVLCSPFTTHSNEY 417
 QY 421 QPNLFNWRMGOONNGASSSGSFTWGNMYPFGGPRLCAGSELAKEMAVFTIHLY 480
 Db 418 EPEKFMPSRFEERHAP-----YTFIPFGGVCTRCPEWERSKQTLILFHYEV 467
 QY 481 LKFMW--ELADDOPFAFPVDP-NGLPYR 509
 Db 468 KTFSGYIPLDPEKVLGNPVPPLRANGFAIKL 499

RESULT 15

US-10-142-231-91
 ; Sequence 91, Application US/10142231
 ; Publication No. US20030077796A1
 ; GENERAL INFORMATION:
 ; APPLICANT: Croteau, Rodney et al.
 ; TITLE OF INVENTION: CYTOCHROME P450 OXYGENASES AND THEIR USES
 ; FILE REFERENCE: 62773
 ; CURRENT APPLICATION NUMBER: US/10/142,231
 ; PRIOR FILING DATE: 2002-05-08
 ; PRIOR APPLICATION NUMBER: 60/165,250
 ; PRIOR FILING DATE: 1999-11-12
 ; NUMBER OF SEQ ID NOS: 95
 ; SOFTWARE: Patentln Ver. 2.1
 ; SEQ ID NO 91
 ; LENGTH: 498
 ; TYPE: PRF
 ; ORGANISM: *Taxus cuspidata*
 US-10-142-231-91

Query Match 20.6%; Score 553.5; DB 9; Length 498;
 Best Local Similarity 29.6%; Pred. No. 6,4e-37;

Matches 140; Conservative 101; Mismatches 191; Indels 41; Gaps 10;

QY 17 SLILSLFLILKRRNRKTRFNLPGKSGMPFLGETIGLYKPYATTLGDMDQVSKYG 76
 Db 30 SALAGILLILFLFRKSHSLKLPQKIGIPFGISFTFLRLRNSLSLOFPDERYKRG 89
 QY 77 KIYRNLGCEPTIVSADAGLNFILQNGRLFECSYPSISGIGILGKWSMLVLYGMDHMD 136
 Db 90 LVEKTSLSLGHPTVYLCGPAGNRLILSNNEKLVQMSWPAQPMKMGENSEVATRRGDHYM 149
 QY 137 RSLNLFSLHARLTILK---DVERHTFLVDSMOONSIFSADQAKKFTFNMAKHIM 193
 Db 150 KSLAGFGPGALQSYIGKMTIELQSH---INERKKGDEYVNVPLVRELFNLSAILLF 206
 QY 194 SMPDGEETBOLKEEYTFPKGVYSAFLNPGTAVHKALOSRTIILFIRKMEERKDI 253
 Db 207 NI-YDKQODRLHLLETTYMGAVCIPLDPGTNYRKALRLKIDKVLSSL 265
 QY 254 KEOOEDEEVKTE--DEAMSKSDHVRKQRTDDLLGVNLKHSNSTEQDILILSLFLA 311
 Db 266 SLDRQOPRISLCSLSEMTKG-----LPH---PMDEILDNFSILHA 306
 QY 312 GHETSSVALAIFLQACPAVEELREHELEIARAKKEGSELNWDYKMDPTOCYI 371
 Db 307 SYDPTTSPMALILFLKSSNTECEYVVOLEIL-SHREKEE-ELTWKDLKAMYTWQVA 364
 QY 372 NETLRLGNVYRFLHKKALKDVRKYDIPSGMKVLPVISAHLNDRYDQPNLFNWRMGO 431
 Db 365 QETLRMFPLFGSFRKALTDIHYDITLIPKMRVLCSPFTTHSNEY 417
 QY 432 QONNGASSSGSFTWGNMYPFGGPRLCAGSELAKEMAVFTIHLYLAKFN 484
 Db 424 QEGKHVAP-----YTFIPFGGQORSCVGNERSKMEILFLVHHFVTFSS 466

Wed Jun 11 10:32:35 2003

Search completed: June 10, 2003, 22:03:35
Job time : 33 secs

us-09-502-426b-2.rapb

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: June 10, 2003, 21:57:28 ; Search time 162 Seconds

(without alignments)
2041.655 Million cell updates/sec

Title: US-09-502-426B-2

Perfect score: 2681

Sequence: 1 MFEHHTLPLLLPSLLS.....FAFPVDFPNCPLTRVSRIL 513

Scoring table:

BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 4569144 seqs, 644733110 residues

Total number of hits satisfying chosen parameters: 4569144

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

```
1: /cgn2_6/ptodata/1/paa/PCRNUS.COMB.pep.*
2: /cgn2_6/ptodata/1/paa/US06.COMB.pep.*
3: /cgn2_6/ptodata/1/paa/US07.COMB.pep.*
4: /cgn2_6/ptodata/1/paa/US08.COMB.pep.*
5: /cgn2_6/ptodata/1/paa/US08.COMB.pep.*
6: /cgn2_6/ptodata/1/paa/US08.COMB.pep.*
7: /cgn2_6/ptodata/1/paa/US08.COMB.pep.*
8: /cgn2_6/ptodata/1/paa/US08.COMB.pep.*
9: /cgn2_6/ptodata/1/paa/US08.COMB.pep.*
10: /cgn2_6/ptodata/1/paa/US08.COMB.pep.*
11: /cgn2_6/ptodata/1/paa/US08.COMB.pep.*
12: /cgn2_6/ptodata/1/paa/US08.COMB.pep.*
13: /cgn2_6/ptodata/1/paa/US08.COMB.pep.*
14: /cgn2_6/ptodata/1/paa/US09.COMB.pep.*
15: /cgn2_6/ptodata/1/paa/US09.COMB.pep.*
16: /cgn2_6/ptodata/1/paa/US09.COMB.pep.*
17: /cgn2_6/ptodata/1/paa/US09.COMB.pep.*
18: /cgn2_6/ptodata/1/paa/US09.COMB.pep.*
19: /cgn2_6/ptodata/1/paa/US09.COMB.pep.*
20: /cgn2_6/ptodata/1/paa/US09.COMB.pep.*
21: /cgn2_6/ptodata/1/paa/US09.COMB.pep.*
22: /cgn2_6/ptodata/1/paa/US09.COMB.pep.*
23: /cgn2_6/ptodata/1/paa/US09.COMB.pep.*
24: /cgn2_6/ptodata/1/paa/US10.COMB.pep.*
25: /cgn2_6/ptodata/1/paa/US10.COMB.pep.*
26: /cgn2_6/ptodata/1/paa/US10.COMB.pep.*
27: /cgn2_6/ptodata/1/paa/US10.COMB.pep.*
```

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	2681	100.0	513	1	PCT-US00-03820-2
2	2681	100.0	513	19	US-09-502-426-2
3	2677	99.9	513	21	US-09-791-537-67246
4	1500	55.9	392	18	US-09-464-535-36
5	1016	37.9	484	21	US-09-708-427-51512
6	1016	37.9	485	21	US-09-708-427-51511

7	1016	37.9	542	21	US-09-708-427-51510	Sequence 51510, A
8	978.5	36.5	472	17	US-09-306-844-2	Sequence 2, Appl1
9	978.5	36.5	472	19	US-09-513-996A-55847	Sequence 55847, A
10	978.5	36.5	472	19	US-09-513-996A-55847	Sequence 56459, A
11	978.5	36.5	472	21	US-09-791-537-145750	Sequence 145750, A
12	978.5	36.5	472	23	US-09-935-625-5227	Sequence 5227, Ap
13	978.5	36.5	472	23	US-09-935-625-5227	Sequence 5227, Ap
14	978.5	36.5	472	23	US-09-935-625-5227	Sequence 5227, Ap
15	978.5	36.5	472	23	US-09-935-625-5227	Sequence 5227, Ap
16	978.5	36.5	472	23	US-09-935-625-5227	Sequence 5227, Ap
17	978.5	36.5	472	23	US-09-935-625-5227	Sequence 5227, Ap
18	978.5	36.5	472	23	US-09-935-625-5227	Sequence 5227, Ap
19	978.5	36.5	472	23	US-09-935-625-5227	Sequence 5227, Ap
20	977.5	36.5	472	23	US-09-935-625-5227	Sequence 5227, Ap
21	946	35.3	444	19	US-09-513-996A-55848	Sequence 55848, A
22	946	35.3	444	19	US-09-513-996A-55848	Sequence 55848, A
23	946	35.3	444	23	US-09-935-625-5228	Sequence 5228, A
24	946	35.3	444	23	US-09-935-625-5228	Sequence 5228, A
25	946	35.3	444	23	US-09-935-625-5228	Sequence 5228, A
26	946	35.3	444	23	US-09-935-625-5228	Sequence 5228, A
27	866	32.3	401	23	US-09-935-625-5228	Sequence 5228, A
28	866	32.3	401	23	US-09-935-625-5228	Sequence 5228, A
29	866	32.3	401	23	US-09-935-625-5228	Sequence 5228, A
30	832	31.0	499	26	US-10-219-999-40933	Sequence 40933, A
31	832	31.0	503	27	US-10-219-999-40933	Sequence 40933, A
32	832	31.0	503	27	US-10-219-999-40933	Sequence 40933, A
33	822	30.7	362	23	US-09-995-917A-1	Sequence 917A, A
34	816.5	30.5	486	25	US-10-142-231-92	Sequence 17051, A
35	808	30.0	153	21	PCT-US99-2285B-1487	Sequence 20716, A
36	805	30.0	153	21	PCT-US99-2285B-1487	Sequence 3487, Ap
37	792	29.5	518	27	US-10-219-999-45776	Sequence 45776, A
38	792	29.5	518	27	US-10-219-999-45776	Sequence 45776, A
39	792	29.5	518	27	US-10-219-999-45776	Sequence 45776, A
40	790	28.5	515	26	US-10-219-999-45776	Sequence 45776, A
41	761.5	28.4	464	21	US-09-791-537-17918	Sequence 17918, A
42	754	28.1	489	27	US-10-337-358-381	Sequence 381, App
43	754	28.1	490	26	US-10-337-358-381	Sequence 62803, A
44	754	28.1	490	27	US-10-337-358-381	Sequence 31870, A
45	739.5	27.6	497	26	US-10-219-999-54439	Sequence 54439, A

ALIGNMENTS

```
RESULT 1
PCT-US00-03820-2
; Sequence 2, Application PC/TUS0003820
; GENERAL INFORMATION:
; APPLICANT: THE ARIZONA BOARD OF REGENTS ON BEHALF OF THE UNIVERSITY OF ARIZONA
; TITLE OF INVENTION: DMP4 POLYNUCLEOTIDES, POLYPEPTIDES AND USES THEREOF
; FILE REFERENCE: 2225-0001.40
; CURRENT APPLICATION NUMBER: PCT/US00/03820
; EARLIER FILING DATE: 2000-02-11
; EARLIER APPLICATION NUMBER: 60/119, 657
; EARLIER FILING DATE: 1999-02-11
; EARLIER APPLICATION NUMBER: 60/119, 658
; NUMBER OF SEQ ID NOS: 18
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 2
; LENGTH: 513
; TYPE: PRT
; ORGANISM: Arabidopsis sp.
PCT-US00-03820-2
```

```
Query Match 100.0%; Score 2681; DB 1; Length 513;
Best Local Similarity 100.0%; Pred. No. 3.4e-231;
Matches 513; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 MFEHHTLPLLLPSLLSLLFLKRRNRKTRNPPGSGMPFLGTTGYLKPYT 60
DB 1 MFEHHTLPLLLPSLLSLLFLKRRNRKTRNPPGSGMPFLGTTGYLKPYT 60
```

Qy	61	ATTLEDPMOONHXSXGKIYRSMLFPEPTIVSADAGINREITLONEGRLFECSVPRIJGIL	120
Dd	61	ATTLEDPMOONHXSXGKIYRSMLFPEPTIVSADAGINREITLONEGRLFECSVPRIJGIL	120
Qy	121	GKWSMLVLVGDMHRDMRSISLNFLSHARLRTLLDKVERHNTLFVLDSMOONSIFSAQDEA	180
Dd	121	GKWSMLVLVGDMHRDMRSISLNFLSHARLRTLLDKVERHNTLFVLDSMOONSIFSAQDEA	180
Qy	181	KKFFNMLAKHIMSMDPGEETEOJKEVYTFMKGVSAPLMPCGAYHKAQOSRATYILK	240
Dd	181	KKFFNMLAKHIMSMDPGEETEOJKEVYTFMKGVSAPLMPCGAYHKAQOSRATYILK	240
Qy	241	FIERKMEBRKLDIKEDDEEVEVKTEDAEAMSKSDHVAKQRTDDDLGLGVNLKHSNLSTEQ	300
Dd	241	FIERKMEBRKLDIKEDDEEVEVKTEDAEAMSKSDHVAKQRTDDDLGLGVNLKHSNLSTEQ	300
Qy	301	ILDLILSLFLFAGHEHSSVAIALAIFFLQACPRAAVEELREBHETIARAKKELGSEELNMD	360
Dd	301	ILDLILSLFLFAGHEHSSVAIALAIFFLQACPRAAVEELREBHETIARAKKELGSEELNMD	360
Qy	361	YKKMDFTOCVINETIRLGNVYRFLRRKALKDVRYKGYDIPSGMKVLPVYSAAVHLNDSRD	420
Dd	361	YKKMDFTOCVINETIRLGNVYRFLRRKALKDVRYKGYDIPSGMKVLPVYSAAVHLNDSRD	420
Qy	421	QPNLFNPMRMOOONNGASSSGSGSFTWGNNTMPFGGGRCLCAGSELALEMAVFIHLIV	480
Dd	421	QPNLFNPMRMOOONNGASSSGSGSFTWGNNTMPFGGGRCLCAGSELALEMAVFIHLIV	480
Qy	481	LKFNNMELADEDOPFAFPVDFPNGJPIRISRTL	513
Dd	481	LKFNNMELADEDOPFAFPVDFPNGJPIRISRTL	513

```

RESULT 2
US-09-502-426-2
;
; Sequence 2, Application US/09502426
;
; GENERAL INFORMATION:
;
; APPLICANT: AZPIROZ, Ricardo
;
; APPLICANT: CHOE, Sunghwa
;
; APPLICANT: FELDMANN, Kenneth
;
; TITLE OF INVENTION: DMFA POLYNUCLEOTIDES, POLYPEPTIDES AND USES THEREOF
;
; FILE REFERENCE: 2225-0001
;
; CURRENT APPLICATION NUMBER: US/09/502,426
;
; CURRENT FILING DATE: 2000-02-11
;
; EARLIER APPLICATION NUMBER: 60/119,657
;
; EARLIER FILING DATE: 1999-02-11
;
; EARLIER APPLICATION NUMBER: 60/119,658
;
; EARLIER FILING DATE: 1999-02-11
;
; NUMBER OF SEQ ID NOS: 18
;
; SOFTWARE: PatentIn Ver. 2.0
;
; SEQ ID NO 2
;
; LENGTH: 513
;
; TYPE: PR1
;
; ORGANISM: Arabidopsis sp.
;
; US-09-502-426-2

```

Query Match	100.0%;	Score 2681;	DB 19;	Length 513;
Best Local Similarity	100.0%;	Pred. No. 3.4e-231;		
Matches 513;	Conservative 0;	Mismatches 0;	Indels 0;	Gaps 0

QY	1	MEFEHNHLLP	LLP	SL	SL	FL	IL	IL	KL	RR	RR	K	R	R	N	P	P	G	S	G	P	P	G	E	T	I	G	L	K	P	T	60					
Db	1	MEFEHNHLLP	LLP	SL	SL	FL	IL	IL	KL	RR	RR	K	R	R	N	P	P	G	S	G	P	P	G	E	T	I	G	L	K	P	T	60					
QY	61	ATTGGDEP	QOQ	HV	S	K	G	K	I	Y	R	S	N	L	F	G	E	P	T	V	S	A	D	A	G	I	N	P	T	L	N	E	G	R	L	120	
Db	61	ATTGGDEP	QOQ	HV	S	K	G	K	I	Y	R	S	N	L	F	G	E	P	T	V	S	A	D	A	G	I	N	P	T	L	N	E	G	R	L	120	
QY	61	ATTGGDEP	QOQ	HV	S	K	G	K	I	Y	R	S	N	L	F	G	E	P	T	V	S	A	D	A	G	I	N	P	T	L	N	E	G	R	L	120	
Db	61	ATTGGDEP	QOQ	HV	S	K	G	K	I	Y	R	S	N	L	F	G	E	P	T	V	S	A	D	A	G	I	N	P	T	L	N	E	G	R	L	120	
QY	121	GKWSMLV	LV	G	M	H	R	D	M	R	S	I	S	N	F	L	S	H	A	R	L	T	I	L	L	D	V	E	N	T	L	F	V	L	D	S	180
Db	121	GKWSMLV	LV	G	M	H	R	D	M	R	S	I	S	N	F	L	S	H	A	R	L	T	I	L	L	D	V	E	N	T	L	F	V	L	D	S	180
QY	121	GKWSMLV	LV	G	M	H	R	D	M	R	S	I	S	N	F	L	S	H	A	R	L	T	I	L	L	D	V	E	N	T	L	F	V	L	D	S	180
Db	121	GKWSMLV	LV	G	M	H	R	D	M	R	S	I	S	N	F	L	S	H	A	R	L	T	I	L	L	D	V	E	N	T	L	F	V	L	D	S	180
QY	181	KKFFPNL	MA	K	H	I	M	S	D	P	E	E	E	T	E	O	L	K	E	V	T	F	M	G	V	A	S	A	P	L	N	P	G	T	A	240	
Db	181	KKFFPNL	MA	K	H	I	M	S	D	P	E	E	E	T	E	O	L	K	E	V	T	F	M	G	V	A	S	A	P	L	N	P	G	T	A	240	

Db	161	KKFTNLAKHMSMDPGEETEDQLKKEVTFPKGVASPLNPGAIYHALOSSRATILK	240
QY	241	FIERRMEERKLDIKREDEDEEEBEVKTEDAEAMSKSDHVQRORTDDDLGWLKHSNLSTEQ	300
Db	241	FIERKMEERKLDIKREDEDEEEBEVKTEDAEAMSKSDHVQRORTDDDLGWLKHSNLSTEQ	300
QY	301	ILDLILSLFLFNGHEHSSVAIALAIFPLQCPRAVEELREHLEIAARKKELGSELNWD	360
Db	301	ILDLILSLFLFNGHEHSSVAIALAIFPLQCPRAVEELREHLEIAARKKELGSELNWD	360
QY	361	YKMDFTOCVINETRLGNNVRFELHRKALKDVRKYKDIPISGWKVLPVISAVHLDNSRYD	420
Db	361	YKMDFTOCVINETRLGNNVRFELHRKALKDVRKYKDIPISGWKVLPVISAVHLDNSRYD	420
QY	421	OPNLFNPRMOOQNNNGASSSGSGSFSTWNNNTYFPGGGRLLCAGSELAKLEMAVFIHLV	480
Db	421	OPNLFNPRMOOQNNNGASSSGSGSFSTWNNNTYFPGGGRLLCAGSELAKLEMAVFIHLV	480
QY	481	LKFNNELAEDDOPFAFPFVDFPNGGJPIRYSRLL 513	
Db	481	LKFNNELAEDDOPFAFPFVDFPNGGJPIRYSRLL 513	

```

RESULT 3
US-09-791-537-67246
: Sequence 67246, Application US/09791537
: GENERAL INFORMATION:
: APPLICANT: Bionomix, Inc.
: APPLICANT: Debe, Derek
: APPLICANT: Danzer, Joseph
: TITLE OF INVENTION: THREE DIMENSIONAL STRUCTURES OF PROTEIN FAMILIES AND FAMILY
: TITLE OF INVENTION: METHODS OF USE THEREOF
: FILE REFERENCE: 261/210
: CURRENT APPLICATION NUMBER: US/09/791,537
: CURRENT FILING DATE: 2001-02-22
: NUMBER OF SEQ ID NOS: 153055
: SOFTWARE: PatentIn version 3.0
: SEQ ID NO 67246
: LENGTH: 513
: TYPE: PRT
: ORGANISM: Arabidopsis thaliana
US-09-791-537-67246

```

Query Match	99.98;	Score 2677;	DB 21;	Length 513;
Best Local Similarity	99.88;	Pred. No. 7.8e-231;		
Matches 512; Conservative	1;	Mismatches .0;	Indels 0;	Gaps 0;

QY	1	MFEENHLLPLLLPLSLLSLLFTLLKRRRRKTRFNLPPCKSGSPRLGETTGKLPPT	60
Db	1	MEFEHHLLPLLLPLSLLSLLFTLLKRRRRKTRFNLPPCKSGSPRLGETTGKLPPT	60
QY	61	ATTGDFMQOHVSKGYIKYRSNLFGEPTVTSADAGLNRETLONEGTLFECSPYSIGIL	120
Db	61	ATTGDFMQOHVSKGYIKYRSNLFGEPTVTSADAGLNRETLONEGTLFECSPYSIGIL	120
QY	121	GKWSMLVLVGDMHRDMRSISLNFSLSHARLTLLDVERHTLVLDSDWOONSIFSADDEA	180
Db	121	GKWSMLVLVGDMHRDMRSISLNFSLSHARLTLLDVERHTLVLDSDWOONSIFSADDEA	180
QY	181	KKFFPNMLAKIIMSMDPGEETEOUKKEYVTFMKGVSAPLMLPGTAIYKALOSRATTLK	240
Db	181	KKFFPNMLAKIIMSMDPGEETEOUKKEYVTFMKGVSAPLMLPGTAIYKALOSRATTLK	240
QY	241	FIEERMEERKLDIKEEDDEEEVKTDEDAEMSKSHVRRQRTDDLLGVLNHSNSTEQ	300
Db	241	FIEERMEERKLDIKEEDDEEEVKTDEDAEMSKSHVRRQRTDDLLGVLNHSNSTEQ	300
QY	301	ILDLILSLFPGHETSVAIALAIFPLDACPAVABELREENLEIARAKKELGESSELNMD	360
Db	301	ILDLILSLFPGHETSVAIALAIFPLDACPAVABELREENLEIARAKKELGESSELNMD	360
QY	361	YKKADFTQCVINETLRLGNVREFLRKALKDVRKGYDIPSGMKVLPIVISAHLNDSRYD	420

```

Db      361 YKKMDFQCVINETLRLGNVRFHRRALKDVKRYKGYDIPSGMKVLPVISAHLNDRYD 420
      421 QPNLFNPRMROQNNNGASSSGSFSFTGNNYMPFGGPRLCASSELAKLEMAVFIIHLV 480
      421 QPNLFNPRMROQNNNGASSSGSFSFTGNNYMPFGGPRLCASSELAKLEMAVFIIHLV 480
      481 LKFNWELAEDDQPAFPFVDFPENGPIRVSRIL 513
      481 LKFNWELAEDDQPAFPFVDFPENGPIRVSRIL 513

```

RESULT 4

```

US-09-464-535-36
; Sequence 36, Application US/09464535
; GENERAL INFORMATION:
; APPLICANT: Famodu, Omolayo O.
; APPLICANT: Cahoon, Rebecca E.
; APPLICANT: Sakai, Hajime
; APPLICANT: McGonigle, Brian
; APPLICANT: Rafalski, J. Antoni
; TITLE OF INVENTION: STEROL BIOSYNTHETIC ENZYMES
; FILE REFERENCE: BB1306 US NA
; CURRENT APPLICATION NUMBER: US/09/464,535
; EARLIER FILING DATE: 1999-12-15
; EARLIER APPLICATION NUMBER: 60/112,555
; NUMBER OF SEQ ID NOS: 44
; SOFTWARE: Microsoft Office 97
; SEQ ID NO 36
; LENGTH: 392
; TYPE: PRT
; ORGANISM: Oryza sativa
US-09-464-535-36

```

```

Query Match      55.9%; Score 1500; DB 18; Length 392;
Best Local Similarity 70.3%; Pred. No. 2.9e-125;
Matches 289; Conservative 40; Mismatches 54; Indels 28; Gaps 2;

```

```

QY      102 QNEGRLEPCSTPRSTIGGILGKWSMLVYGDHNRMRKRSISLNFSLHARKLTLLKDVREHT 161
      3 QNEGRLEPCSTPRSTIGGILGKWSMLVYGDHNRMRKRSISLNFSLHARKLTLLKDVREHT 62
Db      162 LFLVDSMOQNSIFSAODEAKKFTFNLMKNTMSMDPGEETEOLKKEVTMKGVSAPL 221
      63 LFLVDSMOQNSIFSAODEAKKFTFNLMKNTMSMDPGEETEOLKKEVTMKGVSAPL 122
QY      222 NLPGTAYHAKALQSRAITLFTIERKMEERKLDIKEEDQEEBEVKTEDAEAMSKSDHVRKOR 281
      123 NLPGTAYHAKALQSRAITLFTIERKMEERKLDIKEEDQEEBEVKTEDAEAMSKSDHVRKOR 163
QY      282 TDDDLGKVLKHSNISTQIIDLILSLFAGHETSSVAIALAIFLQACPKVAEELREH 341
      164 --DDLGLWALQNSINKEQIIDLILSLFAGHETSSVAIALAIFLQACPKVAEELREH 221
QY      342 LEIARAKKEGESELMNDYKKMDPTOCVINETLRLGNVRFHRRALKDVKRYKGYDIPS 401
      222 LGIARQRLRECKISWEDYKEMFTQCVINETLRLGNVRFHRRALKDVKRYKGYDIPS 281
Db      402 GKMVLPIVISAHLNDRYDQPNLFNPRMROQNNNGASSSGSFSFTGNNYMPFGGPRLC 461
      282 GKMVLPIVISAHLNDRYDQPNLFNPRMROQNNNGASSSGSFSFTGNNYMPFGGPRLC 334
QY      462 CAGSELAKLEMAVFIIHLVLFKFNWELAEDDQPAFPFVDFPENGPIRVSRIL 512
      335 CAGSELAKLEMAVFIIHLVLFKFNWELAEDDQPAFPFVDFPENGPIRVSRIL 385
Db

```

```

RESULT 5
US-09-708-427-51512
; Sequence 51512, Application US/09708427
; GENERAL INFORMATION:
; APPLICANT: N. ALEXANDROV et al.

```

```

; TITLE OF INVENTION: SEQUENCE-DETERMINED DNA FRAGMENTS AND CORRESPONDING POLYPEPTID
; FILE REFERENCE: 2750-1243P
; CURRENT APPLICATION NUMBER: US/09/708,427
; CURRENT FILING DATE: 2000-11-09
; NUMBER OF SEQ ID NOS: 85364
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 51512
; LENGTH: 484
; TYPE: PRT
; ORGANISM: Zea mays subsp. mays
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: 1..484
; OTHER INFORMATION: Xaa is any amino acid
; NAME/KEY: misc_feature
; LOCATION: 1..484
; OTHER INFORMATION: Ceres Seq. ID 1927049
US-09-708-427-51512

```

```

Query Match      37.9%; Score 1016; DB 21; Length 484;
Best Local Similarity 41.4%; Pred. No. 1.2e-81;
Matches 213; Conservative 102; Mismatches 143; Indels 56; Gaps 11;

```

```

QY      7 HTLLPL--LLPLSLILL--FILLRRNRKRTFNPFGKSGMPLGETIGYLKPYTAT 62
      9 HXLAALCTLLASLTLLVNLHFLPL--LNPKA----PRSPFGMLGETLRLFTLRPNASN 62
Db      63 TLGDFMOQHVSKYKIRYRNLFGPTTVSADAGINRFTLONEGRLEPCSTPRSTIGILGK 122
      63 TLGDFLHDHCSRYGRVFKSHLFCPTVAVSCDQDLNHFILONEGRLEPCSTPRSTIGILGK 122
QY      123 WSMVLVYGDHNRMRKRSISLNFSLHARKLTLLKDVREHTLFLVDSMOQNS-----IFSA 176
      123 WSMVLVYGDHNRMRKRSISLNFSLHARKLTLLKPSYLGDIKIALHVGAMRHSGSGGVYVAF 182
QY      177 QDEAKKFTFNLMKNTMSMDPGEETEOLKKEVTMKGVSAPLNPGTAYHAKALQSR 236
      183 CEARKAFSYIVQVGLDSPEEYPTARILEDPLAFKGLISPLYITPGTPYAKAVARE 242
QY      237 TLIERKMEERKLDIKEEDQEEBEVKTEDAEAMSKSDHVRKORTDDLLGKVSNTL 296
      243 RISSTVGIILKERR-----SAGSNKQ---GDFLDVLLSNNEL 277
QY      297 STEQIIDLILSLFAGHETSSVAIALAIFLQACPKVAEELREHLEIARAKKEGESE 355
      278 SDEKVSFVLDSLLGGETYSLLISNVYVFLGQSAQDLDLVLRHDSI--RSNK--GKEEC 334
Db      356 LNMDDYKKMDPTOCVINETLRLGNVRFHRRALKDVKRYKGYDIPSGMKVLPVISAHLN 415
      335 LNSDYKKMEYTOQVINELALRCGNIVAFVHRKALKDVKRYKEXYLIPSGMKVLPVFTVHLN 394
QY      416 NSRYDQPNLFNPRMROQNNNGASSSGSFSFTGNNYMPFGGPRLCASSELAKLEMAVF 475
      395 PSLGDXQOQPCWMEGTSQTS-----KRPFGCGGRLXPSSELAKLEMAVF 443
QY      476 IHHVLKFNWELAEDDQPAFPFVDFPENGPIRV 509
      444 LHHLLNRMRIXDXDIPMAVYXXFORGLPIET 477
Db

```

```

RESULT 6
US-09-708-427-51511
; Sequence 51511, Application US/09708427
; GENERAL INFORMATION:
; APPLICANT: N. ALEXANDROV et al.
; TITLE OF INVENTION: SEQUENCE-DETERMINED DNA FRAGMENTS AND CORRESPONDING POLYPEPTID
; FILE REFERENCE: 2750-1243P
; CURRENT APPLICATION NUMBER: US/09/708,427
; CURRENT FILING DATE: 2000-11-09
; NUMBER OF SEQ ID NOS: 85364
; SOFTWARE: PatentIn version 3.1

```

```

; SEQ ID NO 51511
; LENGTH: 485
; TYPE: PRT
; ORGANISM: Zea mays subsp. mays
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: 1..485
; OTHER INFORMATION: Xaa is any amino acid
; NAME/KEY: misc feature
; LOCATION: 1..485
; OTHER INFORMATION: Ceres Seq. ID 1927048
US-09-708-427-51511

```

```

Query Match      37.9%; Score 1016; DB 21; Length 485;
Best Local Similarity 41.4%; Pred. No. 1.2e-81;
Matches 213; Conservative 102; Mismatches 143; Indels 56; Gaps 11;

```

```

QY 7 HTLLPL--LLPSLLSL--FLILKRRNRKTRFNLPRGSGMPLGEGTIGLKRYAT 62
   | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 10 HXLAALGTLTLLASLTLLVNLHFLPL--LNPKA---PRGSGMPLGEGTIGLKRYAT 63
   | | | | | | | | | | | | | | | | | | | | | | | | | | | |
QY 63 TLGDMOQHVSKYKIRSNLFEPTIVSADAGLNRLFIQNGRLFECSYPSISGIGLK 122
   | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 64 TLGFLFDHCHSRVGRVFKSHLFCPTIVSCDQDLNHLFIQNGRLFECSYPSISGIGLK 123
   | | | | | | | | | | | | | | | | | | | | | | | | | | | |
QY 123 WSMVLVGDMDHRMRSISLNFSLHARLRTILKDVERRHTLVLDWQONS-----IFSA 176
   | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 124 SSMVLVVGEDHRRRLNLALALVSTKLKPSYLGDIKIALHVVGAMRRHSGSGVAVAF 183
   | | | | | | | | | | | | | | | | | | | | | | | | | | | |
QY 177 ODEAKKFTFNMAKHIMSDPGEETEOQLKRYTFMKGVSAPLNLPGTAYHKAQSORA 236
   | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 184 CEARKRAFVSIVKOVGLSPPEPVARTLEDFLAFMKGLISFPLYTPGTAKAVARARE 243
   | | | | | | | | | | | | | | | | | | | | | | | | | | | |
QY 237 TLKFIERKMEERKLDIKEEDQEEVEKTEDEAEMSKSDHYRKORTDDLLGVNLKHSNL 296
   | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 244 RISSTVKGILKER-----SAGSNMKQ---GDFLDVLLSSNEL 278
   | | | | | | | | | | | | | | | | | | | | | | | | | | | |
QY 297 STEQILDILSLFAGHETSSVALAIFLQACPKAVEELREHLEIARAKKEGSE- 355
   | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 279 SDEKVSFVLDLGLGYETTSLLISMVYVFLQSOADDLVLRREHDSI--RSNK--GKEEC 335
   | | | | | | | | | | | | | | | | | | | | | | | | | | | |
QY 356 LNMDDYKMDFTQCVINETLRLGNVRFHRRKALDVRKGYDIPSGKVPVLSAVHLD 415
   | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 336 LISEDYKMEITQOVINEALRCGNIVKFKHRRKALDVKYKEYLLPSGMKVLVPTFAVHLN 395
   | | | | | | | | | | | | | | | | | | | | | | | | | | | |
QY 416 NSRYDQPNLFNPMWMOQONNGASSGGSFSTWGNMNYMPFGGPRLCAGSELAKEMAVF 475
   | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 396 PSLHGDXXQFQPCRMEGTSQGS-----KRFPPFGGPRLXGSELAKVETAF 444
   | | | | | | | | | | | | | | | | | | | | | | | | | | | |
QY 476 IHHVLKFNWELAEDDQPPAFPFVDFPNCGLPIRV 509
   | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 445 LHHLLXNLRWRKIXDDIPMAYPYXXFQRLPIEI 478
   | | | | | | | | | | | | | | | | | | | | | | | | | | | |

```

RESULT 7

```

US-09-708-427-51510
; Sequence 51510, Application US/09708427
; GENERAL INFORMATION:
; APPLICANT: N. ALEXANDROV et al.
; TITLE OF INVENTION: SEQUENCE-DETERMINED DNA FRAGMENTS AND CORRESPONDING POLYPEPTIDES
; TITLE OF INVENTION: THEREBY
; FILE REFERENCE: 2750-1243P
; CURRENT APPLICATION NUMBER: US/09/708.427
; EARLIER FILING DATE: 2000-11-09
; NUMBER OF SEQ ID NOS: 85364
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 51510
; LENGTH: 542
; TYPE: PRT
; ORGANISM: Zea mays subsp. mays
; NAME/KEY: misc feature
; LOCATION: 1..542

```

```

; OTHER INFORMATION: Xaa is any amino acid
; NAME/KEY: misc feature
; LOCATION: 1..542
; OTHER INFORMATION: Ceres Seq. ID 1927047
US-09-708-427-51510

```

```

Query Match      37.9%; Score 1016; DB 21; Length 542;
Best Local Similarity 41.4%; Pred. No. 1.4e-81;
Matches 213; Conservative 102; Mismatches 143; Indels 56; Gaps 11;

```

```

QY 7 HTLLPL--LLPSLLSL--FLILKRRNRKTRFNLPRGSGMPLGEGTIGLKRYAT 62
   | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 67 HXLAALGTLTLLASLTLLVNLHFLPL--LNPKA---PRGSGMPLGEGTIGLKRYAT 120
   | | | | | | | | | | | | | | | | | | | | | | | | | | | |
QY 63 TLGDMOQHVSKYKIRSNLFEPTIVSADAGLNRLFIQNGRLFECSYPSISGIGLK 122
   | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 121 TLGFLFDHCHSRVGRVFKSHLFCPTIVSCDQDLNHLFIQNGRLFECSYPSISGIGLK 180
   | | | | | | | | | | | | | | | | | | | | | | | | | | | |
QY 123 WSMVLVGDMDHRMRSISLNFSLHARLRTILKDVERRHTLVLDWQONS-----IFSA 176
   | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 181 SSMVLVVGEDHRRRLNLALALVSTKLKPSYLGDIKIALHVVGAMRRHSGSGVAVAF 240
   | | | | | | | | | | | | | | | | | | | | | | | | | | | |
QY 177 ODEAKKFTFNMAKHIMSDPGEETEOQLKRYTFMKGVSAPLNLPGTAYHKAQSORA 236
   | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 241 CEARKRAFVSIVKOVGLSPPEPVARTLEDFLAFMKGLISFPLYTPGTAKAVARARE 300
   | | | | | | | | | | | | | | | | | | | | | | | | | | | |
QY 237 TLKFIERKMEERKLDIKEEDQEEVEKTEDEAEMSKSDHYRKORTDDLLGVNLKHSNL 296
   | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 301 RISSTVKGILKER-----SAGSNMKQ---GDFLDVLLSSNEL 335
   | | | | | | | | | | | | | | | | | | | | | | | | | | | |
QY 297 STEQILDILSLFAGHETSSVALAIFLQACPKAVEELREHLEIARAKKEGSE- 355
   | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 336 SDEKVSFVLDLGLGYETTSLLISMVYVFLQSOADDLVLRREHDSI--RSNK--GKEEC 392
   | | | | | | | | | | | | | | | | | | | | | | | | | | | |
QY 356 LNMDDYKMDFTQCVINETLRLGNVRFHRRKALDVRKGYDIPSGKVPVLSAVHLD 415
   | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 393 LISEDYKMEITQOVINEALRCGNIVKFKHRRKALDVKYKEYLLPSGMKVLVPTFAVHLN 452
   | | | | | | | | | | | | | | | | | | | | | | | | | | | |
QY 416 NSRYDQPNLFNPMWMOQONNGASSGGSFSTWGNMNYMPFGGPRLCAGSELAKEMAVF 475
   | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 453 PSLHGDXXQFQPCRMEGTSQGS-----KRFPPFGGPRLXGSELAKVETAF 501
   | | | | | | | | | | | | | | | | | | | | | | | | | | | |
QY 476 IHHVLKFNWELAEDDQPPAFPFVDFPNCGLPIRV 509
   | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 502 LHHLLXNLRWRKIXDDIPMAYPYXXFQRLPIEI 535
   | | | | | | | | | | | | | | | | | | | | | | | | | | | |

```

RESULT 8

```

US-09-306-844-2
; Sequence 2, Application US/09306844
; GENERAL INFORMATION:
; APPLICANT: Koncez, Jasda
; APPLICANT: Mathur, Jaldeep
; APPLICANT: Szekeres, Miklos
; APPLICANT: Altman, Thomas
; TITLE OF INVENTION: NUCLEIC ACID MOLECULES ENCODING CYTOCHROME P450-TYPE
; TITLE OF INVENTION: PROTEINS INVOLVED IN THE BRASSINOSTEROID SYNTHESIS IN
; FILE REFERENCE: 147-186
; CURRENT APPLICATION NUMBER: US/09/306.844
; EARLIER FILING DATE: 1999-05-06
; EARLIER FILING DATE: 1996-03-27
; NUMBER OF SEQ ID NOS: 12
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 2
; LENGTH: 472
; TYPE: PRT
; ORGANISM: Arabidopsis thaliana
US-09-306-844-2

```

```

Query Match      36.5%; Score 978.5; DB 17; Length 472;
Best Local Similarity 41.9%; Pred. No. 2.7e-78;

```


QY 252 DIKEEDDEEVEKTEDEAEMSKSDHVRKQRTDDLLGAVLKHSNSTEQIIDLILSLFA 311
DB 233 VVAKRREDEEGAE-----RKMDLAALLAADGFSDEIDVFLVALVA 278
QY 312 GHETSSVAIALAIFLQACRAVEELREHELEIARAKKELGSELNMDYKMDTQCVI 371
DB 279 GYETSTIMTLAVFLTEPLALQKKEHEKIRAKMSD--SYSEMSDYKSMPTQCV 336
QY 372 NETLRGNAVFLHRLKALKDVRKGYDIPSGMKVLPVISAHLNDSRYDQNLFPNRMQ 431
DB 337 NETLRVANITIGVFRRAATDVEIKGYKIPKGMKVFSFRAYHLDNHFKAARTFPMRMQ 396
QY 432 QONNGASSSGSSTWGN--YMPGGPRLCAGSELAKLEMAVFIHHLVLFKNMELAD 490
DB 397 -----SNSVTTGPSNVFTPPGGPRLCPGELARVALSVFLHRLVTFGSVPAQ 446
QY 491 DQPAFPVDFPNGLPIRVS 511
DB 447 DKLVFFPTTRQKRPPIFKR 467

RESULT 11
US-09-791-537-145750
; Sequence 145750, Application US/09791537
; GENERAL INFORMATION:
; APPLICANT: Biomolix, Inc.
; APPLICANT: Debe, Derek
; APPLICANT: Danzer, Joseph
; TITLE OF INVENTION: THREE DIMENSIONAL STRUCTURES OF PROTEIN FAMILIES AND FAMILY MEMBE
; FILE REFERENCE: 261/210
; CURRENT APPLICATION NUMBER: US/09/791,537
; CURRENT FILING DATE: 2001-02-22
; NUMBER OF SEQ ID NOS: 153055
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 145750
; LENGTH: 472
; TYPE: PRT
; ORGANISM: Arabidopsis thaliana
US-09-791-537-145750

Query Match 36.5%; Score 978.5; DB 21; Length 472;
Best Local Similarity 41.9%; Pred. No. 2,7e-78;
Matches 210; Conservative 79; Mismatches 171; Indels 41; Gaps 8;

QY 12 LLLPSSLILFLILKRRNKRTRFNLPGKSGMPFGETIGYLPKPYATTLGDFMOOH 71
DB 7 LLLSSIAAGFL--LLRRTYRRMGILPGSLGLPIGETFQILGAVKTEPNPEPTIDER 63
QY 72 VSKYKIRYRNLFGEPTIVSADAGLNRFILONEGRLEFECSPRSIGILGKWSMLVYGD 131
DB 64 VARGSVFMTHLFGEPITFSADPETNRFLVONEGKLEFCSPASICNMLGKSHLLMKGS 123
QY 132 MHRDMRSISLNFSLHARLRTILKDVRRHTLFVLDMSQONSIFSAODEAKKTFENLMAKH 191
DB 124 LHKRMSLJMSFANSIIKDHMLDIDRLVRFNDSMSRYLL--MEBAKKTIFELVYQ 181
QY 192 IMSNDPGEETEOLKKEVTFMGVVSAPLNPCTAYHKAQSRATILKFERKMEERKL 251
DB 182 LMSFDPG--EWSSESLRKEVLYIEGFSPLPLFSTTYRKAIOAR-----RKVAEALT 232
QY 252 DIKEEDDEEVEKTEDEAEMSKSDHVRKQRTDDLLGAVLKHSNSTEQIIDLILSLFA 311
DB 233 VVAKRREDEEGAE-----RKMDLAALLAADGFSDEIDVFLVALVA 278
QY 312 GHETSSVAIALAIFLQACRAVEELREHELEIARAKKELGSELNMDYKMDTQCVI 371
DB 279 GYETSTIMTLAVFLTEPLALQKKEHEKIRAKMSD--SYSEMSDYKSMPTQCV 336
QY 372 NETLRGNAVFLHRLKALKDVRKGYDIPSGMKVLPVISAHLNDSRYDQNLFPNRMQ 431
DB 337 NETLRVANITIGVFRRAATDVEIKGYKIPKGMKVFSFRAYHLDNHFKAARTFPMRMQ 396

QY 432 QONNGASSSGSSTWGN--YMPGGPRLCAGSELAKLEMAVFIHHLVLFKNMELAD 490
DB 397 -----SNSVTTGPSNVFTPPGGPRLCPGELARVALSVFLHRLVTFGSVPAQ 446
QY 491 DQPAFPVDFPNGLPIRVS 511
DB 447 DKLVFFPTTRQKRPPIFKR 467

RESULT 12
US-09-935-625-5227
; Sequence 5227, Application US/09935625
; GENERAL INFORMATION:
; APPLICANT: N. ALEXANDROV et al.
; TITLE OF INVENTION: POLYNUCLEOTIDES, POLYPEPTIDES, CELLS, AND METHODS THEREOF C-A
; FILE REFERENCE: 2750-1481P
; CURRENT APPLICATION NUMBER: US/09/935,625
; CURRENT FILING DATE: 2001-08-24
; NUMBER OF SEQ ID NOS: 33136
; SEQ ID NO 5227
; LENGTH: 472
; TYPE: PRT
; ORGANISM: Arabidopsis thaliana
; FEATURE:
; NAME/KEY: peptide
; LOCATION: 1..472
; OTHER INFORMATION: Ceres Seq. ID no. 2114147
US-09-935-625-5227

Query Match 36.5%; Score 978.5; DB 23; Length 472;
Best Local Similarity 41.9%; Pred. No. 2,7e-78;
Matches 210; Conservative 79; Mismatches 171; Indels 41; Gaps 8;

QY 12 LLLPSSLILFLILKRRNKRTRFNLPGKSGMPFGETIGYLPKPYATTLGDFMOOH 71
DB 7 LLLSSIAAGFL--LLRRTYRRMGILPGSLGLPIGETFQILGAVKTEPNPEPTIDER 63
QY 72 VSKYKIRYRNLFGEPTIVSADAGLNRFILONEGRLEFECSPRSIGILGKWSMLVYGD 131
DB 64 VARGSVFMTHLFGEPITFSADPETNRFLVONEGKLEFCSPASICNMLGKSHLLMKGS 123
QY 132 MHRDMRSISLNFSLHARLRTILKDVRRHTLFVLDMSQONSIFSAODEAKKTFENLMAKH 191
DB 124 LHKRMSLJMSFANSIIKDHMLDIDRLVRFNDSMSRYLL--MEBAKKTIFELVYQ 181
QY 192 IMSNDPGEETEOLKKEVTFMGVVSAPLNPCTAYHKAQSRATILKFERKMEERKL 251
DB 182 LMSFDPG--EWSSESLRKEVLYIEGFSPLPLFSTTYRKAIOAR-----RKVAEALT 232
QY 252 DIKEEDDEEVEKTEDEAEMSKSDHVRKQRTDDLLGAVLKHSNSTEQIIDLILSLFA 311
DB 233 VVAKRREDEEGAE-----RKMDLAALLAADGFSDEIDVFLVALVA 278
QY 312 GHETSSVAIALAIFLQACRAVEELREHELEIARAKKELGSELNMDYKMDTQCVI 371
DB 279 GYETSTIMTLAVFLTEPLALQKKEHEKIRAKMSD--SYSEMSDYKSMPTQCV 336
QY 372 NETLRGNAVFLHRLKALKDVRKGYDIPSGMKVLPVISAHLNDSRYDQNLFPNRMQ 431
DB 337 NETLRVANITIGVFRRAATDVEIKGYKIPKGMKVFSFRAYHLDNHFKAARTFPMRMQ 396
QY 432 QONNGASSSGSSTWGN--YMPGGPRLCAGSELAKLEMAVFIHHLVLFKNMELAD 490
DB 397 -----SNSVTTGPSNVFTPPGGPRLCPGELARVALSVFLHRLVTFGSVPAQ 446
QY 491 DQPAFPVDFPNGLPIRVS 511
DB 447 DKLVFFPTTRQKRPPIFKR 467

RESULT 13
US-09-935-625-5735

```
; Sequence 5735, Application US/09935625
; GENERAL INFORMATION:
; APPLICANT: N. ALEXANDROV et al.
; TITLE OF INVENTION: POLYNUCLEOTIDES, POLYPEPTIDES, CELLS, AND METHODS THEREOF CAPABLE
; FILE REFERENCE: 2750-1481P
; CURRENT APPLICATION NUMBER: US/09/935,625
; CURRENT FILING DATE: 2001-08-24
; NUMBER OF SEQ ID NOS: 33136
; SEQ ID NO 5735
; LENGTH: 472
; TYPE: PRT
; ORGANISM: Arabidopsis thaliana
; FEATURE:
; NAME/KEY: peptide
; LOCATION: 1..472
; OTHER INFORMATION: Ceres Seq. ID no. 3058391
US-09-935-625-5735
```

```
Query Match
Best Local Similarity 36.5%; Score 978.5; DB 23; Length 472;
Matches 210; Conservative 79; Mismatches 171; Indels 41; Gaps 8;
```

```
QY 12 LLLPSLLSLFLILKRRNRKTRFNPFGKSGMPFLGFTIGYIKPYTATLGDPMQOH 71
DB 7 LLLSSIAAGFL---LLLRTRRRMGILPPSGILGLIGETPOLIGAYKTENPEPFIDER 63
QY 72 VSKYKITYRSNLFCEPTIVSADAGLNRTILONEGRLEFCSYPRISGIGILGKSMVLVGD 131
DB 64 VARGSVPMTHLFREPTIFSDPETNRPVLONEGRLEFCSYPASICNLGKSHSLMGKS 123
QY 132 MHRDMRSISLNFSLHARLRTILKDVREHTLFLVDSMQONSIFSAODEAKKFTNLMK 191
DB 124 LHKRHSLSLMSFANSSTIKDHLMDIDLVRFNLSWSRVLL--MEAKKITFELTVKQ 181
QY 192 IMSDPEEETEOLKREYVTPMKGVASAPLNPCTAYHKAQSATIIKFERKMEERKL 251
DB 182 LMSFDPG-EMSESLRKEYLVIEGFSLPLPFSITYRKAIQAR-----RKVAEALT 232
QY 252 DIKEEDOEBEVKTEDEAKMSKSDHVRKORTDDDLGVKLKHSNISTQIIDLILSLFA 311
DB 233 VVVMKRRREEEGAE-----RKMDLAALLAADGFSDEIVDFVALVLA 278
QY 312 GHETSSVAIALAIFLOACPRAVEELREHELEIARAKEGESELMNDYKKMDTQCVI 371
DB 279 GYETTSTIMTLAVKFLTETPLALQKKEHEKIRAKMSD--SYSEMSDYKSMPTQCV 336
QY 372 NETLRIGNVRFLLHRAKLDVRYKGYDIPSGMKVLPVISAHLNDSRYDQPLFNPWRMQ 431
DB 337 NETLRVANIIGVFRAMTDEIKGYKIPKGMKVFSSFRAYHLDPNHFKDARTFNPWRMQ 396
QY 432 QONNGASSSGSSTGNN--YMPFGGPRLCAGSELAKLEMAVTHHLVLFKNMELAD 490
DB 397 -----SNVTTGPSNVFTFPGGPRLCPEYELARVALSVFLHRLVTFGSVWPAEQ 446
QY 491 DQPAFPPVDPNGLPIRVSR 511
DB 447 DKLVFFPTTRTKRYPFVKR 467
```

```
RESULT 14
US-09-935-625-7739
; Sequence 7739, Application US/09935625
; GENERAL INFORMATION:
; APPLICANT: N. ALEXANDROV et al.
; TITLE OF INVENTION: POLYNUCLEOTIDES, POLYPEPTIDES, CELLS, AND METHODS THEREOF CAPABLE
; FILE REFERENCE: 2750-1481P
; CURRENT APPLICATION NUMBER: US/09/935,625
; CURRENT FILING DATE: 2001-08-24
; NUMBER OF SEQ ID NOS: 33136
; SEQ ID NO 7739
; LENGTH: 472
```

```
; TYPE: PRT
; ORGANISM: Arabidopsis thaliana
; FEATURE:
; NAME/KEY: peptide
; LOCATION: 1..472
; OTHER INFORMATION: Ceres Seq. ID no. 2114147
US-09-935-625-7739
```

```
Query Match
Best Local Similarity 36.5%; Score 978.5; DB 23; Length 472;
Matches 210; Conservative 79; Mismatches 171; Indels 41; Gaps 8;
```

```
QY 12 LLLPSLLSLFLILKRRNRKTRFNPFGKSGMPFLGFTIGYIKPYTATLGDPMQOH 71
DB 7 LLLSSIAAGFL---LLLRTRRRMGILPPSGILGLIGETPOLIGAYKTENPEPFIDER 63
QY 72 VSKYKITYRSNLFCEPTIVSADAGLNRTILONEGRLEFCSYPRISGIGILGKSMVLVGD 131
DB 64 VARGSVPMTHLFREPTIFSDPETNRPVLONEGRLEFCSYPASICNLGKSHSLMGKS 123
QY 132 MHRDMRSISLNFSLHARLRTILKDVREHTLFLVDSMQONSIFSAODEAKKFTNLMK 191
DB 124 LHKRHSLSLMSFANSSTIKDHLMDIDLVRFNLSWSRVLL--MEAKKITFELTVKQ 181
QY 192 IMSDPEEETEOLKREYVTPMKGVASAPLNPCTAYHKAQSATIIKFERKMEERKL 251
DB 182 LMSFDPG-EMSESLRKEYLVIEGFSLPLPFSITYRKAIQAR-----RKVAEALT 232
QY 252 DIKEEDOEBEVKTEDEAKMSKSDHVRKORTDDDLGVKLKHSNISTQIIDLILSLFA 311
DB 233 VVVMKRRREEEGAE-----RKMDLAALLAADGFSDEIVDFVALVLA 278
QY 312 GHETSSVAIALAIFLOACPRAVEELREHELEIARAKEGESELMNDYKKMDTQCVI 371
DB 279 GYETTSTIMTLAVKFLTETPLALQKKEHEKIRAKMSD--SYSEMSDYKSMPTQCV 336
QY 372 NETLRIGNVRFLLHRAKLDVRYKGYDIPSGMKVLPVISAHLNDSRYDQPLFNPWRMQ 431
DB 337 NETLRVANIIGVFRAMTDEIKGYKIPKGMKVFSSFRAYHLDPNHFKDARTFNPWRMQ 396
QY 432 QONNGASSSGSSTGNN--YMPFGGPRLCAGSELAKLEMAVTHHLVLFKNMELAD 490
DB 397 -----SNVTTGPSNVFTFPGGPRLCPEYELARVALSVFLHRLVTFGSVWPAEQ 446
QY 491 DQPAFPPVDPNGLPIRVSR 511
DB 447 DKLVFFPTTRTKRYPFVKR 467
```

```
RESULT 15
US-09-935-625-9349
; Sequence 9349, Application US/09935625
; GENERAL INFORMATION:
; APPLICANT: N. ALEXANDROV et al.
; TITLE OF INVENTION: POLYNUCLEOTIDES, POLYPEPTIDES, CELLS, AND METHODS THEREOF CAPA
; FILE REFERENCE: 2750-1481P
; CURRENT APPLICATION NUMBER: US/09/935,625
; CURRENT FILING DATE: 2001-08-24
; NUMBER OF SEQ ID NOS: 33136
; SEQ ID NO 9349
; LENGTH: 472
; TYPE: PRT
; ORGANISM: Arabidopsis thaliana
; FEATURE:
; NAME/KEY: peptide
; LOCATION: 1..472
; OTHER INFORMATION: Ceres Seq. ID no. 3058391
US-09-935-625-9349
```

```
Query Match
Best Local Similarity 36.5%; Score 978.5; DB 23; Length 472;
Matches 210; Conservative 79; Mismatches 171; Indels 41; Gaps 8;
```

```
QY 12 LLLPBLLSLLFLILIKRRNRKTRNLPFGKSGWPLFGTIGYLPKPYATTLGDFMOQH 71
Db 7 LLLSSIAAGFL--LLLRTRRYRRMGLPGLPGSLGLPGLGETFQILGAYKTENPEPFIDER 63
QY 72 VSKYGIYRSNLGEPPIVSADAGLNRFILQNEGRLEECSPRSIGILGKMSLVYGD 131
Db 64 VARYGSVFMTLHGEPTIESADPETNRPFVLQNEGKLEECSPASICNLGKHSLLMKGS 123
QY 132 MHRDMRSISLNFSLHARLRTILKDYERHTLFYLDSSWQONSIFSADDEAKKFTFNLMKH 191
Db 124 LHKRMSLIMSFSANSLIKDHLMDIDRLVRFNLDSSSVLL--MEAKKITFELTVKQ 181
QY 192 IMSMDGEEETEOLKEEYVTEMKVVSAPLNLPGTAVHKAQSRATILKFIERKMEERKL 251
Db 182 LMSFDPG-EMSESLRREYLLVIEGFSLPLPLFSTYRKAIQAR-----RKVAELT 232
QY 252 DIKEODOEEEVKTEDEAEKSKSDHYRKQKTDDDLGVNLKHSNLSTEQILDLILSLFA 311
Db 233 VVVMKRREEEGAE-----RKDMLAALLAADDGFSDEETVDELVALVA 278
QY 312 GHETSSVAIALAIFLQACPKAVEELREHLEIARAKKESELMNDYKKMDFTQCVI 371
Db 279 GYETSTIMTLAKFLTEPPLAQAQKEHEKIRAKSD--SYLEMSDYKSMPTQCV 336
QY 372 NETLRGNVVRFLHRAKLDVRYKYDIPSGWKVLPVISAVHLDNSRYDQPNLENPWRQ 431
Db 337 NETLRVANIIGYVFRAMTDVEIKYKIPKMKVFSSEFRAVHLDPNHFKDARTENPWRQ 396
QY 432 QONNGASSSGSSEFSTWGN--VMPFGGPRLCAGSELAKEMAVFIHHLVLFKNWELAE 490
Db 397 -----SNSVTGPSNVFTPPGGPRLCPGELARVALSVFLHRLVTGFSWPAEQ 446
QY 491 DQFAFPFVDFPGLPIRVSR 511
Db 447 DKLVFFPTRTQKRYPIFVR 467
```

Search completed: June 10, 2003, 22:02:53
job time : 163 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: June 10, 2003, 21:58:58 ; Search time 58 Seconds

(Without alignments)
2004.717 Million cell updates/sec

Title: US-09-502-426b-2

Perfect score: 2681 1 MFETHTLPLPLPLSLLS.....FAFPVDFPGLPIRVSRL 513

Scoring table:

BLISSUM62
Gapop 10.0 , Gapext 0.5

Searched: 1050051 seqs, 226654218 residues

Total number of hits satisfying chosen parameters: 1050051

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%

Maximum Match 100%
Listing first 45 summaries

Database :

Pending_Patents_AA.New:*
1: /cgn2_6/pdata/2/paa/PCF_NEW_COMB.pep:*
2: /cgn2_6/pdata/2/paa/US06_NEW_COMB.pep:*
3: /cgn2_6/pdata/2/paa/US07_NEW_COMB.pep:*
4: /cgn2_6/pdata/2/paa/US08_NEW_COMB.pep:*
5: /cgn2_6/pdata/2/paa/US09_NEW_COMB.pep:*
6: /cgn2_6/pdata/2/paa/US10_NEW_COMB.pep:*
7: /cgn2_6/pdata/2/paa/US60_NEW_COMB.pep:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	2681	100.0	513	US-09-502-426a-2	Sequence 2, Appl
2	2681	100.0	513	US-09-502-426b-2	Sequence 2, Appl
3	1770	66.0	506	US-10-395-463-2	Sequence 2, Appl
4	1166.5	43.5	300	US-10-424-599-282744	Sequence 282744,
5	1164	43.4	317	US-10-425-114-72982	Sequence 72982, A
6	1001	37.3	488	US-10-424-599-178693	Sequence 178693,
7	978.5	36.5	472	US-09-502-426a-19	Sequence 19, Appl
8	978.5	36.5	472	US-09-502-426b-19	Sequence 19, Appl
9	856.5	31.9	501	US-10-395-463-24	Sequence 24, Appl
10	849	31.7	505	US-09-502-426a-25	Sequence 25, Appl
11	849	31.7	505	US-09-502-426b-25	Sequence 25, Appl
12	832	31.0	503	US-10-310-154-384	Sequence 384, App
13	816.5	30.5	486	US-10-356-153-92	Sequence 92, App
14	792	29.5	518	US-10-425-114-59005	Sequence 383, App
15	790	29.5	515	US-10-425-114-59005	Sequence 383, App
16	761.5	28.4	464	US-09-502-426a-20	Sequence 20, Appl
17	761.5	28.4	464	US-09-502-426b-20	Sequence 20, Appl
18	754	28.1	489	US-10-310-154-381	Sequence 381, App
19	748	27.9	469	US-10-395-463-30	Sequence 30, App
20	739.5	27.6	497	US-10-425-114-64364	Sequence 64364, A
21	719.5	26.8	504	US-10-424-599-236550	Sequence 236550,
22	642	23.9	453	US-10-259-194A-156	Sequence 156, App
23	641	23.9	294	US-10-425-114-48987	Sequence 48987, A
24	634	23.6	470	US-10-310-154-386	Sequence 386, App
25	622	23.2	495	US-10-425-114-47304	Sequence 47304, A
26	614.5	22.9	496	US-10-424-599-271221	Sequence 271221,

27	602.5	22.5	499	US-10-310-154-382	Sequence 382, App
28	597.5	22.3	485	US-10-356-153-88	Sequence 88, Appl
29	596.5	22.2	497	US-10-356-153-64	Sequence 64, Appl
30	596	22.2	474	US-10-424-599-275182	Sequence 275182,
31	592	22.1	466	US-10-356-153-90	Sequence 90, Appl
32	588	21.9	509	US-10-356-153-67	Sequence 67, Appl
33	585	21.8	507	US-10-356-153-65	Sequence 65, Appl
34	583	21.7	484	US-10-356-153-57	Sequence 57, Appl
35	571	21.3	501	US-10-356-153-58	Sequence 58, Appl
36	569.5	21.2	493	US-10-356-153-89	Sequence 89, Appl
37	568.5	21.2	474	US-10-424-599-248036	Sequence 248036,
38	568.5	21.2	500	US-10-356-153-59	Sequence 59, Appl
39	564	21.0	484	US-10-356-153-56	Sequence 56, Appl
40	564	21.0	500	US-10-356-153-61	Sequence 61, Appl
41	559	20.9	298	US-10-424-599-205981	Sequence 205981,
42	558.5	20.8	483	US-10-356-153-62	Sequence 62, Appl
43	558.5	20.8	512	US-10-356-153-87	Sequence 87, Appl
44	553.5	20.6	498	US-10-356-153-91	Sequence 91, Appl
45	545	20.3	280	US-10-425-114-57231	Sequence 57231, A

ALIGNMENTS

RESULT 1
US-09-502-426a-2
Sequence 2, Application US/09502426A
GENERAL INFORMATION:
APPLICANT: AZPIROZ, Ricardo
APPLICANT: AZPIROZ, Ricardo
APPLICANT: CHO, Sungwa
APPLICANT: FELDMANN, Kenneth
TITLE OF INVENTION: DWF POLYNUCLEOTIDES, POLYPEPTIDES AND USES THEREOF
FILE REFERENCE: 2225-0001
CURRENT APPLICATION NUMBER: US/09/502,426A
CURRENT FILING DATE: 2000-02-11
PRIOR APPLICATION NUMBER: 60/119,657
PRIOR FILING DATE: 1999-02-11
PRIOR APPLICATION NUMBER: 60/119,658
PRIOR FILING DATE: 1999-02-11
NUMBER OF SEQ ID NOS: 25
SOFTWARE: FastSeq for Windows Version 4.0
SEQ ID NO 2
LENGTH: 513
TYPE: PRT
ORGANISM: Arabidopsis sp.
US-09-502-426a-2

Query Match 100.0%; Score 2681; DB 5; Length 513;
Best Local Similarity 100.0%; Pred. No. 8.2e-25;
Matches 513; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
1 MFETHTLPLPLPLSLLSLTLFLILKRRNRKTRFNLPPKSGMPLGRTGYLKYPT 60
1 MFETHTLPLPLPLSLLSLTLFLILKRRNRKTRFNLPPKSGMPLGRTGYLKYPT 60
61 ATTLDPMQOHSKXGKIYRSLRGEPTIVSDAGLNRIILONEGRLEFCSPRIGTL 120
61 ATTLDPMQOHSKXGKIYRSLRGEPTIVSDAGLNRIILONEGRLEFCSPRIGTL 120
121 GKMSLVLVGDHMRDRSISLNLFLSHARLRTILKDVREHTFLVDSMOONSIFSAODEA 180
121 GKMSLVLVGDHMRDRSISLNLFLSHARLRTILKDVREHTFLVDSMOONSIFSAODEA 180
161 KFTFNLMAKHTMSDPEEETBOLKKEYVTFMKGVSAPIPLPTAYHKAQSRATILK 240
161 KFTFNLMAKHTMSDPEEETBOLKKEYVTFMKGVSAPIPLPTAYHKAQSRATILK 240
241 FIERMERKRLDIKEEDDEEVEKEDAEKSKSDHVKORDDLLGWLKHSNLSTEO 300
241 FIERMERKRLDIKEEDDEEVEKEDAEKSKSDHVKORDDLLGWLKHSNLSTEO 300
301 ILDLISLLFAGHETSSVAIALATFFLOACRAVEELREHLEIETARAKEGESELANMD 360

US-10-424-599-282744
Sequence 282744, Application US/10424599
GENERAL INFORMATION:
APPLICANT: La Rosa Thomas J
APPLICANT: Kovalic David K
APPLICANT: Zhou Yihua
APPLICANT: Cao Yongwei
TITLE OF INVENTION: Soy Nucleic Acid Molecules and Other Molecules Associated with
TITLE OF INVENTION: Plants and Uses Thereof for Plant Improvement
FILE REFERENCE: 38-21(53223)B
CURRENT APPLICATION NUMBER: US/10/424,599
CURRENT FILING DATE: 2003-04-28
NUMBER OF SEQ ID NOS: 285684
SEQ ID NO 282744
LENGTH: 300
TYPE: PRT
ORGANISM: glycine max
FEATURE:
OTHER INFORMATION: Clone ID: PAT_MRT3847_97339C.1.pep
US-10-424-599-282744

[illegible]

```

RESULT 5
US-10-425-114-72982
: Sequence 72982, Application US/10425114
: GENERAL INFORMATION:
: APPLICANT: Liu, Jingdong
: APPLICANT: Zhou, Yihua
: APPLICANT: Kovalic, David K.
: APPLICANT: Screen, Steven E
: APPLICANT: Tabaska, Jack E
: APPLICANT: Cao, Yongwei
: TITLE OF INVENTION: Nucleic Acid Molecules and Other Molecules Associated With
: TITLE OF INVENTION: Plants and Uses Thereof for Plant Improvement
: FILE REFERENCE: 38-21(5313)B
: CURRENT APPLICATION NUMBER: US/10-425,114
: CURRENT FILING DATE: 2003-04-28
: NUMBER OF SEQ ID NOS: 73128
: SEQ ID NO 72982
: LENGTH: 317
: TYPE: PRT
: ORGANISM: Zea mays
: FEATURE:
: OTHER INFORMATION: Clone ID: UC-ZMR0B73080G04_F11.pep
US-10-425-114-72982

```

Db 251 KESTTEKK-----NDMLGALLASGYHFSDEITVPMALLVAG 289
Qy 313 HETSSVAIALAIFLQACPKAVEELREHELEIARAKKELGSELMNDYKKMDTOCVIN 372
Db 290 YETSTIMTLAKFLTEPLALAOKEHDOI-RAKSCPEAPLEMDYDKMATQCVN 348
Qy 373 ETLRLGNVVRFLHRAKALKDVRKYGDIPSGMKVLPVISAHLNDSRYDQPNLFNPMWQ 432
Db 349 ETLRVANIGAFRRAMTDINIKGYTIPKGRVVASFVAHLNPDHFKDARTFPMRW-Q 407
Qy 433 QONNGASSSGSFTWGNVWPEGGPRLCAGSELAKLEMAVFIHHLVLEKFNWELAEDDQ 492
Db 408 SNSASSP-----CNVYTPFGGPRLCGYELARVALSVFLHRTVTRYSWPEAEDEK 459
Qy 493 PFAPFVDFPNGLPPIRVSR 511
Db 460 LVFEPTTRTQKRYPIYKR 478

RESULT 7

US-09-502-426a-19
; Sequence 19, Application US/09502426A
; GENERAL INFORMATION:
; APPLICANT: AZPIROZ, Ricardo
; APPLICANT: CHOE, Sungbwa
; APPLICANT: FELDMANN, Kenneth
; TITLE OF INVENTION: DMF4 POLYNUCLEOTIDES, POLYPEPTIDES AND USES THEREOF
; FILE REFERENCE: 2225-0001
; CURRENT APPLICATION NUMBER: US/09/502,426A
; PRIOR FILING DATE: 2000-02-11
; PRIOR APPLICATION NUMBER: 60/119,657
; PRIOR FILING DATE: 1999-02-11
; PRIOR APPLICATION NUMBER: 60/119,658
; NUMBER OF SEQ ID NOS: 25
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 19
; LENGTH: 472
; TYPE: PRT
; ORGANISM: Arabidopsis thaliana
US-09-502-426a-19

Query Match 36.5%; Score 978.5; DB 5; Length 472;

Best Local Similarity 41.9%; Pred. No. 3.2e-86;
Matches 210; Conservative 79; Mismatches 171; Indels 41; Gaps 8;

Qy 12 LLLPSTLSLLFLILKRRNRKTRFNLPGKSGWPFLEGTIGYLRKYATTATLGDPMOQH 71
Db 7 LLLLSIAAGFL---LLRRTYRRMGLPFGSLGLPLIGETPQLIGAYKTENPEPFI DER 63
Qy 72 VSKYGIKIRSNLFGEPITVSADAGLNRFILONEGRLEFCSPRSIGIGILGKMSMLVYGD 131
Db 64 VARGSVFMTHLFGEPITVSADPETNRVFLQNEGRLEFCSPASICNLGKSHLLMGKS 123
Qy 132 MHRDMSISLNFSLHARLITILKDYERHLLFVLDMSQONSIFSAODEAKKFTFNIMAKH 191
Db 124 LHRMHSLLTMSFANSSTIKDHLMDIDLVRFNLDSSSRVLL--MEBAKKTFFELYKQ 181
Qy 192 IMSNDPGEETEQULKEEYVTFMKGVASAPLNLPGTAVHKALOSRAITILKTERKMEERKL 251
Db 182 LMSFDPG-EMSESLRKEVLYIEGFSPLPLSTYTRKAIQAR-----RKVAEALT 232
Qy 252 DIKEEDOEVEEYKTEDEAKMSKSDHVRKQRTDDLLGVLKHSNLSLEQIIDLILSLFA 311
Db 233 VVVMKRRREEEGAE-----RKMDLAALLAADGFSDEIYDFLVALLYA 278
Qy 312 GHETSSVAIALAIFLQACPKAVEELREHELEIARAKKELGSELMNDYKKMDTOCVI 371
Db 279 GYETSTIMTLAVFLETPLALAOKEHEKIRAMKSD--SYLSMDSYKSMPTQCVV 336
Qy 372 NETLRGNVVRFLHRAKALKDVRKYGDIPSGMKVLPVISAHLNDSRYDQPNLFNPMWQ 431

Db 337 NETLRVANIGVFRAMTDVEIKGYKIPKCKVYSSFRVAHLDPNHFKDARTFNPWRMO 396
Qy 432 QONNGASSSGSFTWGNVWPEGGPRLCAGSELAKLEMAVFIHHLVLEKFNWELAE 490
Db 397 -----SNSVYTPFGSNVFTPPGGGPRLCGYELARVALSVFLHRLVTFGSVPAEQ 446
Qy 491 DQFAFPVDFPNGLPPIRVSR 511
Db 447 DKLVEFPTRTQKRYPIYKR 467

RESULT 8

US-09-502-426b-19
; Sequence 19, Application US/09502426B
; GENERAL INFORMATION:
; APPLICANT: AZPIROZ, Ricardo
; APPLICANT: CHOE, Sungbwa
; APPLICANT: FELDMANN, Kenneth A.
; TITLE OF INVENTION: DMF4 POLYNUCLEOTIDES, POLYPEPTIDES AND USES THEREOF
; FILE REFERENCE: 11696-070001
; CURRENT APPLICATION NUMBER: US/09/502,426B
; PRIOR FILING DATE: 2000-02-11
; PRIOR APPLICATION NUMBER: US 60/119,657
; PRIOR FILING DATE: 1999-02-11
; PRIOR APPLICATION NUMBER: 60/119,658
; NUMBER OF SEQ ID NOS: 30
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 19
; LENGTH: 472
; TYPE: PRT
; ORGANISM: Arabidopsis thaliana
US-09-502-426b-19

Query Match 36.5%; Score 978.5; DB 5; Length 472;

Best Local Similarity 41.9%; Pred. No. 3.2e-86;
Matches 210; Conservative 79; Mismatches 171; Indels 41; Gaps 8;

Qy 12 LLLPSTLSLLFLILKRRNRKTRFNLPGKSGWPFLEGTIGYLRKYATTATLGDPMOQH 71
Db 7 LLLLSIAAGFL---LLRRTYRRMGLPFGSLGLPLIGETPQLIGAYKTENPEPFI DER 63
Qy 72 VSKYGIKIRSNLFGEPITVSADAGLNRFILONEGRLEFCSPRSIGIGILGKMSMLVYGD 131
Db 64 VARGSVFMTHLFGEPITVSADPETNRVFLQNEGRLEFCSPASICNLGKSHLLMGKS 123
Qy 132 MHRDMSISLNFSLHARLITILKDYERHLLFVLDMSQONSIFSAODEAKKFTFNIMAKH 191
Db 124 LHRMHSLLTMSFANSSTIKDHLMDIDLVRFNLDSSSRVLL--MEBAKKTFFELYKQ 181
Qy 192 IMSNDPGEETEQULKEEYVTFMKGVASAPLNLPGTAVHKALOSRAITILKTERKMEERKL 251
Db 182 LMSFDPG-EMSESLRKEVLYIEGFSPLPLSTYTRKAIQAR-----RKVAEALT 232
Qy 252 DIKEEDOEVEEYKTEDEAKMSKSDHVRKQRTDDLLGVLKHSNLSLEQIIDLILSLFA 311
Db 233 VVVMKRRREEEGAE-----RKMDLAALLAADGFSDEIYDFLVALLYA 278
Qy 312 GHETSSVAIALAIFLQACPKAVEELREHELEIARAKKELGSELMNDYKKMDTOCVI 371
Db 279 GYETSTIMTLAVFLETPLALAOKEHEKIRAMKSD--SYLSMDSYKSMPTQCVV 336
Qy 372 NETLRGNVVRFLHRAKALKDVRKYGDIPSGMKVLPVISAHLNDSRYDQPNLFNPMWQ 431
Db 337 NETLRVANIGVFRAMTDVEIKGYKIPKCKVYSSFRVAHLDPNHFKDARTFNPWRMO 396
Qy 432 QONNGASSSGSFTWGNVWPEGGPRLCAGSELAKLEMAVFIHHLVLEKFNWELAE 490
Db 397 -----SNSVYTPFGSNVFTPPGGGPRLCGYELARVALSVFLHRLVTFGSVPAEQ 446
Qy 491 DQFAFPVDFPNGLPPIRVSR 511
Db 447 DKLVEFPTRTQKRYPIYKR 467

RESULT 9

US-10-395-463-24

Sequence 24, Application US/10395463.

GENERAL INFORMATION:

APPLICANT: Tanaka, Hiroshi

APPLICANT: Kayano, Toshaki

APPLICANT: Matsuo, Makoto

APPLICANT: Sakamoto, Tomoaki

APPLICANT: Iwahori, Shuichi

TITLE OF INVENTION: Method of Controlling Character of Monocytyledon by Modification

TITLE OF INVENTION: Overexpression of Cytochrome P450 Monooxygenase Gene Involved in

FILE REFERENCE: 59150-8022

CURRENT APPLICATION NUMBER: US/10/395,463

PRIOR FILING DATE: 2003-03-20

PRIOR FILING DATE: 2002-09-20

NUMBER OF SEQ ID NOS: 38

SOFTWARE: PatentIn version 3.1

SEQ ID NO 24

LENGTH: 501

TYPE: PR

ORGANISM: Oryza sativa

US-10-395-463-24

Query Match

Best Local Similarity 31.9%; Score 856.5; DB 6; Length 501;

Matches 191; Conservative 97; Mismatches 189; Indels 41; Gaps 11;

QY 11 PLLLPSSLILFLILK-----RRNRKTRFNLPGKSGMPLEGITGYLKPY 59
 DB 5 PVLIAAAAVVAVVAVLRLWLLILGSPAGRLGK--RALMPGSGTLPILGETRLISAY 62
 QY 60 TATLGDPMQOHKSKYKIRSNLREPTIVSADAGLNRFLLQNEGRLFECSYPSIGIT 119
 DB 63 KIPNPEPTIDERNVARGGVFTTHVEGERVFSADPFRNLILAAEGRAVHSSYSSIALT 122
 QY 120 LGRMSLVLVGDMHRDMSISLNFSLHARLRTLLDKVDVHNTLFVLDNQNSIPSAOE 179
 DB 123 LGRSLLTRGAHKLHSLTLRLGRP-ASPPILAHIDRLVATYROMEPATYVRLME 181
 QY 180 AKKFTNLMAKHIMSMDPEEETDLKKEVTFMGVVSAPLN---LPGTAYHVALOSR 235
 DB 182 AKKFTNLMVQVLEIEPQ-PTWESIAREYVKLIDGFSIPPLANILPFTYGGALKAR 240
 QY 236 ATILKFERKMEERKLDIKEEDQEEEVKTEDEAEMSKSDHVRKQRTDDLLGWLKHSN 295
 DB 241 KKVAAALREVYKRR-----MEERAKNGSGISGDEGKREKEDMY-----EELL--EAEGGS 288
 QY 236 LSTEQILDILISLIFAGHETSSVAIALAIFFLQCPKAVEELREHLEIFARAKKEGSE 355
 DB 289 FSEEMVDFCISLILVAGYETTSMLMTLAVKFLTEPPLALAELEKHANTRDKGK--KOP 346
 QY 356 LMDMDYKMDFTQCINETLRLGNVVRFLHRAKALDVRKGYDIPSGMKVLPVISAVHLD 415
 DB 347 LEMSYKSMPTQVCINETLRLGNVVRFLHRAKALDVRKGYDIPSGMKVLPVISAVHLD 406
 QY 416 NSRYDQPLNFPMRQOONNGASSSGSFTWGNWMPFGGGRPLCAGSELAKEMAVF 475
 DB 407 NEHYENARTFNPWRQOINNKLONAVGA-----NIFPFGGGRPLCAGSELARVVSIF 459
 QY 476 IHHVLKFNWLEAEDDQPFAPFVDFPGLPIRVSRI 513
 DB 460 LHHVTRFSWETEDRLVFFPTTTLKGYPIINL-RL 496

RESULT 10

US-09-502-426A-25

Sequence 25, Application US/09502426A

GENERAL INFORMATION:

APPLICANT: Azpiroz, Ricardo

APPLICANT: CHOE, Sunghwa

APPLICANT: FELDMANN, Kenneth

TITLE OF INVENTION: DMF4 POLYNUCLEOTIDES, POLYPEPTIDES AND USES THEREOF

FILE REFERENCE: 2225-0001

CURRENT APPLICATION NUMBER: US/09/502,426A

PRIOR FILING DATE: 2000-02-11

PRIOR FILING DATE: 1999-02-11

PRIOR FILING DATE: 1999-02-11

NUMBER OF SEQ ID NOS: 25

SOFTWARE: FastSeq for Windows Version 4.0

SEQ ID NO 25

LENGTH: 575

TYPE: PR

ORGANISM: Artificial Sequence

FEATURE: OTHER INFORMATION: Consensus sequence

FEATURE: NAME/KEY: VARIANT

LOCATION: (1)...(575)

OTHER INFORMATION: Xaa = Any Amino Acid or No Amino Acid

US-09-502-426A-25

Query Match

Best Local Similarity 31.7%; Score 849; DB 5; Length 575;

Matches 215; Conservative 35; Mismatches 242; Indels 38; Gaps 10;

QY 15 LPSILSLILFLILK-----NKRTRFNLPGKSGMPLEGITGYLKPYTATL 64
 DB 39 LLSXALVXKLAAARXXXKXXKXXKXXKXXKXXKXXKXXKXXKXXKXXKXXKXXKXX 98
 QY 65 GPMQOHVSKY---KIYRNLFGPTIVSADAGLNRFLLQNEG-RLFECSYPSIGIT 120
 DB 99 GPMQOHVSKY---KIYRNLFGPTIVSADAGLNRFLLQNEG-RLFECSYPSIGIT 157
 QY 121 GKMMSLVLVGDMHRDMSISLNFSLHARLRTLLDKVDVHNTLFVLDNQNSIPSAOE 176
 DB 158 GKMMSLVLVGDMHRDMSISLNFSLHARLRTLLDKVDVHNTLFVLDNQNSIPSAOE 217
 QY 177 QDEARKFTNLMAKHIMSMDPEEET--EOLKKEVTFMGVVSAPLN--LPGTAYHVALOS 234
 DB 218 XXEXKMPEDXXXKXXKXXKXXKXXKXXKXXKXXKXXKXXKXXKXXKXXKXXKXX 277
 QY 235 RATILKFERKMEERKLDIKEEDQEEEVKTEDEAEMSKSDHVRKQRTDDLLGWLKHSN 291
 DB 278 RAFTXAXLXXXXXXXXXXXXXEXRXXEEXXXXXXXXXXXXXXXXXXXXXDGLLXXX 337
 QY 292 -----KHSNLTSTEQILDILISLIFAGHETSSVAIALAIFFLQCPKAVEELREHLEI 344
 DB 338 XXXXDEXXXLSDXEXIDXXLFLFAGHETSSXXLXAVFLXHPVXEXLREHXA 397
 QY 345 ARAKKEISELMDMDYKMDFTQCINETLRLGNVVRFLHRAKALDVRKGYDIPSGMK 404
 DB 398 XRAKXXKXESLTXDXDKKMYTVCINETLRLAIVGKRRXAKKDVINGIYAIKPKWK 457
 QY 405 VLPVISAVHLDNQNSIPSAOE--EOLKKEVTFMGVVSAPLN--LPGTAYHVALOS 464
 DB 458 VYXSTRVAVHLDNQNSIPSAOE--EOLKKEVTFMGVVSAPLN--LPGTAYHVALOS 517
 QY 465 SELAKEMAVFIIHVLKFNWLEAEDDQPFAPFVDFP-----NGLPIRV 509
 DB 518 KELAKLEKXVFLHRLVQXF-WELAXXND-XXXXLVXFTTXXRDXNDLPIKV 565

RESULT 11

US-09-502-426B-25

Sequence 25, Application US/09502426B

GENERAL INFORMATION:

APPLICANT: Azpiroz, Ricardo

APPLICANT: Choe, Sunghwa

APPLICANT: FELDMANN, Kenneth A.

TITLE OF INVENTION: DMF4 POLYNUCLEOTIDES, POLYPEPTIDES AND USES THEREOF

```

; FILE REFERENCE: 11696-070001
; CURRENT APPLICATION NUMBER: US/09/502.426B
; CURRENT FILING DATE: 2000-02-11
; PRIOR APPLICATION NUMBER: US 60/119,657
; PRIOR FILING DATE: 1999-02-11
; PRIOR APPLICATION NUMBER: US 60/119,658
; PRIOR FILING DATE: 1999-02-11
; NUMBER OF SEQ ID NOS: 30
; SOFTWARE: FASTSEQ for Windows Version 4.0
; SEQ ID NO 25
; LENGTH: 575
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Consensus sequence
; NAME/KEY: VARIANT
; LOCATION: (1)...(575)
; OTHER INFORMATION: Xaa = Any Amino Acid or No Amino Acid
US-09-502-426b-25

```

```

Query Match          31.7%; Score 849; DB 5; Length 575;
Best Local Similarity 40.6%; Pred. No. 1.7e-73;
Matches 215; Conservative 35; Mismatches 242; Indels 38; Gaps 10;

QY 15 LPSLISLLFLILIKR-----NRKTRFNLPGKSGWPLGETIGYLPKYATTL 64
DB 39 LLSXALXVXLXLAARXXXXXXRRKXKXLPCTGMLPYIGETIQLKXXXXXX 98
QY 65 GDFMOQHVSKYK---KTYRSNLGEPITVSADAGLNPIILONEG-RLFECYPSISGIL 120
DB 99 GDFPKKENVXXGXKXXKXIKK-HLSEPTIXSDAELNRKXIKLFCXKXPSXXKXL 157
QY 121 GKSMVLVGDMDHMRMSISLNFSLSHARLRTILKDYERHTLFVLDGK---QONSIFSA 176
DB 158 GKXSLXAXGXGKHEKRMKXILXKXFSXXXLIDHXLPIIDRXKSLXKXKXKXKXKX 217
QY 177 QDEKKTFTNLMKHTMSMDGEEET--EQLKREYVFMKGVSAPLNPGTAYHKALOS 234
DB 218 XHEKKKMTFXXXXXXGXKXKXKXKXKXKXKXKXKXKXKXKXKXKXKXKXKX 277
QY 235 RATLTKFERMEERKIDIKEEDOESEEVKTEDEAEMSKSDHVKORITDDLLGMVL--- 291
DB 278 RAFAKXALXXXXXXXXXXXXXIKEXKXEEKXXXXXXXXXXXXXKXKXKXKXKXKX 337
QY 292 -----KHSNLSQEQILDLILSLFAGHETSVALALAFLOACPRAVELEEHLEI 344
DB 338 XXXXEDXXXXXLSDEXIXDXIXLLFAGHETSSXKXAVFKLEHFDVXKXLEHXA 397
QY 345 ARAKKEGESELMNDODKKKDFTOCVINETIRLGNVYRFLHKKALKDYRKGYDIPSGWK 404
DB 398 XRAKXKXKXKXKXKXKXKXKXKXKXKXKXKXKXKXKXKXKXKXKXKXKXKX 457
QY 405 VLPIYSAVHLDNSRYDDPNLFNPMWMOQNNAGSSGSGSFSTGNVMPMGSGPRICAG 464
DB 458 VXSIRAVHLDPPKYPPEKFNPKKXKXKXKXKXKXKXKXKXKXKXKXKXKXKX 517
QY 465 SELAKLEMAVFIIHLLVLFKFWELAEDDQPAFPFVDP-----NGLPYRV 509
DB 518 KELAKLEMAVFIIHLLVLFKFWELAEDDQPAFPFVDP-----NGLPYRV 565

```

```

; APPLICANT: Banu, G.
; APPLICANT: Bell, Erin
; APPLICANT: Boddupalli, Raghava
; APPLICANT: Deikman, Jill
; APPLICANT: Deng, Molian
; APPLICANT: Dong, Jintuo
; APPLICANT: Duff, Stephen M.
; APPLICANT: Galligan, Meghan M.
; APPLICANT: Hinchey, Brenda S.
; APPLICANT: Huang, Shishieh
; APPLICANT: Johnson, G. Richard
; APPLICANT: Jung, Vincent
; APPLICANT: Kretzmer, Keith A.
; APPLICANT: Laccetti, Lucille B.
; APPLICANT: Lai, Chao-Qiang
; APPLICANT: Lee, Gary
; APPLICANT: Liu, Jie-Yi
; APPLICANT: Liu, Jindong
; APPLICANT: Lu, Bin
; APPLICANT: Luethy, Michael M.
; APPLICANT: Lund, Adrian
; APPLICANT: Madson, Linda L.
; APPLICANT: Malloy, Kathleen A.
; APPLICANT: McKiel, Christine L.
; APPLICANT: Miller, Philip W.
; APPLICANT: Padmavathi, Manchi Kant
; APPLICANT: Parnell, Laurence D.
; APPLICANT: Start, William G.
; APPLICANT: Tennessee, Dan
; APPLICANT: Vidya, K.R.
; APPLICANT: Wang, Haiyuo
; APPLICANT: Xin, Zhanqun
; APPLICANT: Xu, Nanfei
; APPLICANT: Yang, Chunzhi
; APPLICANT: Zeng, Xiaoping
; APPLICANT: Zhang, Qiang
; APPLICANT: Zhao, YaJuan
; APPLICANT: Zhou, Li
; TITLE OF INVENTION: Gene Sequences and Uses Thereof in Plants
; FILE REFERENCE: 38-15(52796)B
; CURRENT APPLICATION NUMBER: US/10/310.154
; PRIOR FILING DATE: 2002-12-04
; PRIOR APPLICATION NUMBER: 60/337,358
; NUMBER OF SEQ ID NOS: 736
; SEQ ID NO 384
; LENGTH: 503
; TYPE: PRT
; ORGANISM: Zea mays
US-10-310-154-384

Query Match          31.0%; Score 832; DB 6; Length 503;
Best Local Similarity 37.2%; Pred. No. 6.2e-72;
Matches 191; Conservative 95; Mismatches 191; Indels 36; Gaps 12;

QY 11 PLLPISLISLL-FLILIKR---NRKTRFNLPGKSGWPLGETIGYLPKYATTLG 65
DB 8 PLFLGAVLAAALAGLWVWRAAGRRSASARLPFGSTGPLIGETIRLLAAYKSPME 67
QY 66 DFMQHVSKYK-ITRSNLGEPITVSADAGLNPIILONEGRLPFCSPRSIGLTKWS 124
DB 68 PFIDERVARRGSGFTTHVGEERTVFSADPAFNRLLLAEGRAVASCSPSSIAITLGRS 127
QY 125 MLVVGDMHMRMSISLNFSLSHARLRTILKDYERHTLFVLDGK---QONSIFSAQ--DEAKK 182
DB 128 LLTSGFAHKLHSLITLRLGRP-ASPLLHIDRLVATYRGHDPGAGMARLLDEKK 186
QY 183 FTFNLMKHTMSMDGEEETEQLKREYVFMKGVSAPLN---LPGTAYHKALQSRTI 238
DB 187 ITFNLVWQVLSIBPG-PWFSVSRREVYKLVDFGFSIFPPASLLPFTVGOALKARKK 245
QY 239 LKTFERMEERKIDIKEEDOESEEVKTEDEAEMSKSDHVKORITDDLLGMVLKHSNST 298

```

```
Db 246 AGALREYVRKMD-----KAEDDCASKKNGEKKDVEELL---EARGSPSV 291
QY 299 EQLDILSLFAGHETSSVALAIFLQACPAVEELREHELEIARAKELDESELNW 358
Db 292 EEWDFGLSLVAYETTSVMTLAVKFLTETPALAQKEHST-RGVYKGOA-LEW 349
QY 359 DDYKMDFTQCVINETRLGNVRFELHRKALDVRKYGDIPSGMKYLPVSAVHLNLSR 418
Db 350 SDYKMPFTQCVISETLRVANLISGFRRANTDHFHGVHVPKCKTIFASPRVHLSLDH 409
QY 419 YDQNLFPNPMWMOQNNGASSGSGSFSTWNNMPGCGPRCLCAGSELAKEMAVFIH 478
Db 410 YENARTDPMRMQDGSKSLQASGASLFT-----PFGGPRCLCPGYELARVYVFLHH 463
QY 479 LVLFENNELAEDDQFAFPVDFPENGPIRYSR 511
Db 464 LVTRFSWEAEEDRIVEFTTRILKGPILLRR 496

RESULT 13
US-10-356-153-92
; Sequence 92, Application US/10356153
; GENERAL INFORMATION:
; APPLICANT: Crotteau, Rodney et al.
; TITLE OF INVENTION: CYTOCHROME P450 OXYGENASES AND THEIR USES
; FILE REFERENCE: 62773
; CURRENT APPLICATION NUMBER: US/10/356,153
; CURRENT FILING DATE: 2003-01-29
; PRIOR APPLICATION NUMBER: US/10/142,231
; PRIOR FILING DATE: 2002-05-08
; PRIOR APPLICATION NUMBER: 60/165,250
; PRIOR FILING DATE: 1999-11-12
; NUMBER OF SEQ ID NOS: 95
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 92
; LENGTH: 486
; TYPE: PRT
; ORGANISM: Taxus cuspidata
US-10-356-153-92

Query Match 30.5%; Score 816.5; DB 6; Length 486;
Best Local Similarity 35.1%; Pred. No. 1,9e-70;
Matches 176; Conservative 97; Mismatches 184; Indels 45; Gaps 8;

QY 19 LSLILFLILKRRNK-----TRFNLPPGKSGMPLIGTIGLKPYATTTGDFWQCHV 72
Db 20 LEICVILLLRSNRQNGSANKPLPPGSAGLPRIGETIFLNRDAKSPGRKKPFDEH 79
QY 73 SKYKTYRSNLFGEPT-TIVSADAGNRFILQNEGRLEFESYRSIGGILGKMSMLVGD 131
Db 80 LKYGPIRCSLGRTRAVVSDPEEFKYYLQNEGRLEFESNALAPPNNLIGKYGSLSAVOGE 139
QY 133 MHRDMSISLNLISHARLTILKDERHTLFLVDSWOONSJFSQODEAKKTPNLMKH 191
Db 140 LQRKHAHVAVNLKHEHTLSSDFMEDIDIFQAGMKRKEEGIPLOHKNOYVLNLMAR 199
QY 192 IMSMPGEETBOLKEEYVTPKQVSAPLNPGVAHKALQSRATTIKFERKMEERL 251
Db 200 LLDLPP-SEMGHITKAFDDFVGAVALSEPLNIPGTYARGIARGLILKRIHKICKERR- 257
QY 252 DIKEDEEEVEYKTEDEAEMSKSDHVRKORTDDLLGWLKLKSNSTQOIIDLISLFA 311
Db 258 -----EHPY-----LRNDLLTLVREGTSDETIADTIIFFEVA 292
QY 312 GHETSSVALAIFLQACPAVEELREHELEIARAKELGSESELNMDYKMDFTQCVI 371
Db 293 GVEISAMAMTFAVKTLAENPRALBELRAEHDLALKAKGK-GNEKLTWNIDYQSMKRVHCVI 351
QY 372 NETTLGNVVRFLHKKALKDVRKYGDIPSGMKYLPVSAVHLNLSRQDPLNLPNWMQ 431
Db 352 NETLRLGATVYLFREAKQDITKADDEVLPKGMTVSVFISAHVHGYCKHYEADKFLPWRMQ 411
QY 432 QONNGASSGSGSFSTWNNMPGCGPRCLCAGSELAKEMAVFIHHLVLFENNELAEDD 491
```

```
Db 412 NE-----GOETLEPCYMPGRCGRGLCPGILHARFEIALFHNFTVKFRWEQLEID 462
QY 492 QPFAFPVDFPENGPIRVS-SRI 512
Db 463 RATYFPLPSTENGPIRILYSRV 484

RESULT 14
US-10-310-154-383
; Sequence 383, Application US/10310154
; GENERAL INFORMATION:
; APPLICANT: Edgerton, Michael D
; APPLICANT: Chomet, Paul S.
; APPLICANT: Adams, Thomas H
; APPLICANT: Ruff, Thomas G.
; APPLICANT: Agarwal, Ameeta K.
; APPLICANT: Ahrens, Jeffrey E.
; APPLICANT: Ball, James A.
; APPLICANT: Bannu, G.
; APPLICANT: Bell, Erin
; APPLICANT: Boddupalli, Raghava
; APPLICANT: Deikman, Jill
; APPLICANT: Dong, Molian
; APPLICANT: Duff, Stephen M.
; APPLICANT: Galligan, Meghan M.
; APPLICANT: Hinchey, Brenda S.
; APPLICANT: Huang, Shihshieh
; APPLICANT: Johnson, G. Richard
; APPLICANT: Jung, Vincent
; APPLICANT: Kretzmer, Keith A
; APPLICANT: Laccetti, Lucille B.
; APPLICANT: Lai, Chao-Qiang
; APPLICANT: Lee, Gary
; APPLICANT: Lin, Jie-Yi
; APPLICANT: Liu, Jindong
; APPLICANT: Lu, Bin
; APPLICANT: Luethy, Michael M.
; APPLICANT: Lund, Adrian
; APPLICANT: Madison, Linda L.
; APPLICANT: Malloy, Kathleen A.
; APPLICANT: McKiel, Christine L.
; APPLICANT: Miller, Philip W.
; APPLICANT: Padmavathi, Manohkanti
; APPLICANT: Parnell, Laurence D.
; APPLICANT: Start, William G.
; APPLICANT: Tennesen, Dan
; APPLICANT: Vidya, K.R.
; APPLICANT: Wang, Haiyun
; APPLICANT: Wang, Zhanguo
; APPLICANT: Xu, Nanrel
; APPLICANT: Yang, Chunzhi
; APPLICANT: Zeng, Xiaoping
; APPLICANT: Zhang, Qiang
; APPLICANT: Zhao, Yajuan
; APPLICANT: Zhou, Li
; TITLE OF INVENTION: Gene Sequences and Uses Thereof in Plants
; FILE REFERENCE: 38-15(52796)B
; CURRENT APPLICATION NUMBER: US/10/310,154
; PRIOR FILING DATE: 2002-12-04
; PRIOR APPLICATION NUMBER: 60/337,358
; PRIOR FILING DATE: 2001-12-04
; NUMBER OF SEQ ID NOS: 736
; SEQ ID NO 383
; LENGTH: 518
; TYPE: PRT
; ORGANISM: Zea mays
US-10-310-154-383

Query Match 29.5%; Score 792; DB 6; Length 518;
Best Local Similarity 33.0%; Pred. No. 5.1e-68;
Matches 160; Conservative 121; Mismatches 162; Indels 42; Gaps 9;
```

```

QY 39 LPBKSMPFLGFTIGYKPYATTLGDMQOHVSKYKIRSNLFGEPTIVSADAGLNR 98
DB 59 LPAGSLWPLLGETLAFIRAYSPRESFVEKRRVLTKYKFSHLMGSPAVYSSDPEVSR 118
QY 99 FILLONEGRLEPCSPRSIGILGKMSMLVYGDHMDRHSISLNFSLARLITILKVE 158
DB 119 AVLOADSAFVPMVPRSLMELMGSSILVGGIQRVHGLAGAFKSPQLKAQVTLDMQ 178
QY 159 RHTLEVLDSW--QOONSI-----FSAODEAKKFTFNMAKHIMSMDPGEETEOLK 207
DB 179 RRVRADDMGRHRHSGDDSGMPRVQNEAKSYFELIVRALIGLEEG-DKQYLRQ 237
QY 208 EYVTPMKGVVAPLNLPGTAYHAKALQSRATILKFERKMEERKIDKEEDEBEVKED 267
DB 238 QFOEFIIAGLISLPYKLPGTOLYRSLKAKRMTKLITIQEKR-----KMMSEG 287
QY 268 EAEKSDHVRKQRTDDLLGWVLKHSN--LSTEQIIDLILSLFAGHETSVAIALAIF 325
DB 288 DEDLRGCTHR-----DMIDVLLGNGNDELSDLEISDNMTDFMIPADSVPLVITLAK 341
QY 326 FLQACPRAVEELREHELEIARAKKELESELNMDYKMDTQCYNITELRLGNVRFILH 385
DB 342 YLSCSPILAQOLEEENNELKROKSGAGET-LEWTDYMSLAFQYVITTELKMGNTINGIM 400
QY 386 RKALKDVRKGYDIPSCMKVLPVISAHLNDSRYDQPNLEPMWMOQOONNGASSSGGSF 445
DB 401 RKAARDVEVGRGHLVPKGMRVLYVFRVAHLDAVHDDPFAFMWRMKERPDVAMSG---- 456
QY 446 STMGNNMPFGGPRCLAGSELAKLEMAVFIHHLVLKPNMELADDOFAFPDFPNCGL 505
DB 457 -----FTPEGGRCLDPCGLDLARLEASIFLHLVITNRM-VAEDVYVNPVYRLKRG 509
QY 506 PIRVS 510
DB 510 PIAVT 514

RESULT 15
US-10-425-114-59005
; Sequence 59005, Application US/10425114
; GENERAL INFORMATION:
; APPLICANT: Liu, Jingdong
; APPLICANT: Zhou, Yihua
; APPLICANT: Kovacic, David K.
; APPLICANT: Screen, Steven E
; APPLICANT: Tabaska, Jack E
; APPLICANT: Cao, Yongwei
; TITLE OF INVENTION: Nucleic Acid Molecules and Other Molecules Associated With
; FILE REFERENCE: 38-21(5313)B
; CURRENT APPLICATION NUMBER: US/10/425,114
; CURRENT FILING DATE: 2003-04-28
; NUMBER OF SEQ ID NOS: 73128
; SEQ ID NO 59005
; LENGTH: 515
; TYPE: PRT
; ORGANISM: Zea mays
; FEATURE:
; OTHER INFORMATION: Clone ID: 700101960_FLI.pep
US-10-425-114-59005

```

Query Match 29.5%; Score 790; DB 6; Length 515;
 Best Local Similarity 33.4%; Pred. No. 7.9e-68;
 Matches 168; Conservative 119; Mismatches 170; Indels 46; Gaps 12;

```

QY 23 LFLILKRR-----NRKTRNLPBKSMPFLGFTIGYKPYATTLGDMQOHVSKYK 76
DB 41 LWWLLSRRAAGSKDERRARLPPGSGFGLVGETIDFVSCAYSSRPEAFVDRRLHIG 100
QY 77 K-IYRSNLFGEPTIVSADAGLNRFLONEGRLEPCSPRSIGILGKMSMLVYGDHMD 135
DB 101 SAVFRSHLFGSAIVYTSDAEVSRLVLSHSDARAFVPMVPRSLTELMGESILLINGSIORR 160

```

```

QY 136 MRSISLNFSLARLITILKDYERHTLFLVDSMOONSI-----FSAODEAKKFTFNMAKH 191
DB 161 VAGIVGAFEFKSPQLKAQYTDQMRLAPRLAMKVCASAPPLRIQDHAKTIVFELVNG 220
QY 192 IMSMDPGEETEOLKKEVYTFMKGVVAPLNLPGTAYHAKALQSR--ATILK-FIERKME 247
DB 221 LIGLEAG--PEMOQLKHOFQEFIVGIMSLPIKLPGRLYRSLQAKRMATLIGIIOEKRR 279
QY 248 ERKIDKEEDEBEVEKTEDEAEMSKSDHVRKQRTDDLLGWVLKHSNLTSTEQIIDLILS 307
DB 260 RRRALLEDGEGEGE-----ACPPRVYIDL--ISGDELTDLEISDNMTD 323
QY 308 LIFAGHETSVAIALAIFELQACPRAVEELREHELEIARAKKELESELNMDYKMDT 367
DB 324 LMIPADSVPLVITLAVKYLSCSPILAQOLEEENMQLRKRTDVGET-LQWTDYMSLSPT 382
QY 368 QCVINETRLGNVRFILHAKALKDVRKGYDIPSCMKVLPVISAHLNDSRYDQPNLEPM 427
DB 383 QHVITETLRMGNTINGIMRKAARDVEVGRGHLIPKGMCVFVYFRSVHLDDKRYDEPYRFP 442
QY 428 WRMQOONNGASSSGGSFSTMGNNMPFGGPRCLAGSELAKLEMAVFIHHLVLKPNMEL 487
DB 443 WRMKKEDT-----STMG--FTPEGGRCLDPCGLDLARLEASIFLHLVTSFRM-V 489
QY 488 AEDDQPFAPFVDFPNCGLPIRVS 510
DB 490 AEEDHIVNPFVYRLKRGMPVRLT 512

```

Search completed: June 10, 2003, 22:04:44
 Job time : 59 secs


```

0Y 12 LLLPSSLLPILLFLLLRNRKTPFNPLPGKSGMFPGETTGYLKLPYATTLIGDMQOH 71
Db 7 LLLLSIAGLU---LLLRTRYRMGCLPGSLGLPILGETTQOLIAVYTENDEPPTDR 63
QY 72 VSKYGIKYSNLFGEPITVSADAGLNRFLLIQNEGRLLFECSPYRIGIGLKGWMLVGD 131
Db 64 VARGSVYMTLHFGEPITFSADPEPNRFLQNEGRKLFECSPASTIONLKGSHLLMKGS 123
QY 132 MHRDMRSTLSNPLSHARLTILKVERHTLVLDWMOONSIFSAODEAKKFTENIMAKH 191
Db 124 LHKRMHSTLTMFSANSSIIKDHMLMDIDRLVRFNLDSWSRVLV--MEBAKKTFFELTVQ 181
QY 192 IMSMDPGEELTOLKKEVYFKGGVVASPLNPGVAHYHALGSRATILKFIKKMEERKL 251
Db 182 IMSDPG-EMWSESKKELTLLVIEGFSLPLPSTTYRAIGAR-----RKVEALT 232
QY 252 DIKEDEDEEEVKTDEDEAEMSKSDHYRKQRTDDLLGWLKHSNLSTEOILLILSLFA 311
Db 233 VYVMKRRREEEGAE-----RKMDMLAALLAADDGFDEELVDPVALLVA 278
QY 312 GHEIYSSAIALAIFFLQACPKAVELREBHLEIARAKKELGSELNWDYKKMDPTQCVI 371
Db 279 GYEETSTMTLAVKFLTETPLALQLKHEHEKIRAMKSD--SYSLEMSDYKSNMFPQCVY 336
QY 372 NETLRIGVAVRFLHKKALKDVRKGYDIPSGKKVLVPIVASIHLNDSRYOQNLFPNWRMQ 431
Db 337 NETLRVAVIIGGVFRAMTDVEIKYKIPKGMKVSSEFRAVHLDHPHEDARTFNWRMQ 396
QY 432 QONNGASSSGSGSEFTMGNN--YMPGGGPRCLACGSELAKLEMAVFTHNHLVLFKNWELEAD 490
Db 397 -----SNSYTTGPSNVFTPPGGGPRCLCPGELMARVALSVFLHRLVLTGSEWVPALQ 446
QY 491 DQPAFPVPVDFPNGLPIVSR 511
I: : : : :

```

Db 450 GGTRECPGAETARIOTAIETIHYETTYKWTOT.KEDRTSEFPSABLYNGEKTOTNR 504


```

Db      5 LIFSFFGLCIFTALLRMNOVYKQANLPPGTMGMFEGETEFLK-----LGPSEM 58
QY      69 QOHVSKYKTYRSMIFGEPTIVSADAGLNREITLONEGRFECSPRSIGILGKMSMVL 128
      59 KNOARVGSFPEKSHILGPTIVSMDSELNRIILVNEAKGLVPGVPSMIDILGKNIAAV 118
QY      129 VGDHHRMRSLNLFSLHARLRTLLKDYERHRLFVLDSDMOONSIFSADAKKFTFNUM 188
      119 NSAHKRYMGALSLISPTMRIDLPKIDEFMRSHLNM-DNKVIDIOEKTNKMAFLSS 177
QY      189 AKHMSMDPGEET---EOLKKEYVTFMKGVYSAPLNPGVYHKAQSRATILKTERK 245
      178 LKQI---AGIESTSLAOEMSEFPNLVGLTSLPINLPNTNHYRGFOARKIIVNLFTL 233
QY      246 MEERKLDIKEDDEEVEEKTEDEAEMSKSDHVRKQRTDDLLGWLAKHS---NLSTEQI 301
      234 IEER-----ASKEOHMDLGYLMEARFRLTIDDEM 266
QY      302 LDLLSLFAGHETSSVALAIFLQACPAVELEREHLEIARAKKEGSELNMDY 361
      267 IDLTITLYSGYEVSTSMAYVYLDHDPKVELEKREHMAIREKKP--EDPIDYNDY 324
QY      362 KKMDFQCVINETLRGNVYRFLRKALKDVRKGYDIPSGKYLPIVSAVHLNDSRYDQ 421
      325 RSMRFTAVILERSRLATTYNGVLRKTTQDMETNGITLIPKGRVYVYTRRLNDYDRLPYD 384
QY      422 PNLFPWRMOQOONASSSGSFSTWGNMYMPFGGPRLCAGSELAKLEMAVFIHLVL 481
      385 PYSFNPWRMMDKS-----LEHONSFLVGGGRGRCGKELGVAELSTFLHYVT 433
QY      482 KFMELAEDDQPFAPFVDPNGLPIRVS 510
      434 KYRWEITGDKLMKFPVEAPNGLIRVS 462

```

RESULT 7

```

T0444
cytochrome P450 - Arabidopsis thaliana
N:Alternate names: protein T18B16.200; protein T5K18.10
C:Species: Arabidopsis thaliana (mouse-ear cross)
C:Date: 30-Apr-1999 #sequence_revision 30-Apr-1999 #text_change 28-Jul-2000
C:Accession: T04444; T05806
R:Bevan, M.; Benes, V.; Reichmann, S.; Borkova, D.; Ansoorge, W.; Bancroft, I.; Mewes, H.W.
submitted to the Protein Sequence Database, April 1998
A:Reference number: Z15359
A:Accession: T04444
A:Molecule type: DNA
A:Residues: 1-457 <BEV>
A:Cross-references: EMBL:AL021687
R:Bevan, M.; Van Der Schueren, J.; Chuang, Y.J.; Voet, M.; Robben, J.; Volckaert, G.; Be
submitted to the Protein Sequence Database, April 1998
A:Reference number: Z15453
A:Accession: T05806
A:Molecule type: DNA
A:Residues: 131-457 <BEV>
A:Cross-references: EMBL:AL022580
A:Experimental source: cultivar Columbia; BAC clone T5K18
C:Genetics:
A:Map position: 4
A:Introns: 67/2; 173/3; 302/3; 358/1; 393/3
A:Note: T18B16.200; T5K18.10
C:Superfamily: Synechocystis cytochrome P450 slr0574; cytochrome P450 homolog
C:Keywords: chromoprotein; electron transfer; heme; iron; monooxygenase; oxidoreductase
F:272-433/Domain: cytochrome P450 homolog <P45>

```

```

Query Match      23.8%; Score 632; DB 2; Length 457;
Best Local Similarity 31.8%; Pred. No. 1.8e-33;
Matches 151; Conservative 92; Mismatches 176; Indels 56; Gaps 12;
QY      13 LLLPSLLSLFLILKRRNRKTRFNLPPKSGMPLGERTIGYIKPTATTLGDPMQOHV 72

```

```

Db      10 LFAGSLFYFLNCLLSQRFPGSSKLLPPLGTMGWPEYGEET---FOLYSDDP-NVFGQSK 65
QY      73 SKYGVIRYNSLGEPTIYSADAGLNREITLONEGRFECSPRSIGILGKMSMVL 132
      66 KRYGVSEFTHVIGCCVAMISPEAKFVLVTKSHLFKPTFPSPAKERMALGKOAIFPHQGY 125
QY      133 HDMSRISLNLFSHARLRTLLKDYERHRLFVLDSDMOONSIFSADAKKFTFNUM 192
      126 HAKLRKLVLRAPMPSIRN-NVPDIESTIADSLRSM-ECOTMINTVOEMKTYFNVALLSI 183
QY      193 MSMDGEET---EOLKKEYVTFMKGVYSAPLNPGVYHKAQSRATILKTERK 249
      184 F---GKDEVLVREDKRCYLLLEGYSMPNLPGLTFHKMKRKELSQIARLTSR 239
QY      250 KLDIKEDDEEVEEKTEDEAEMSKSDHVRKQRTDDLLG-VWLKNSNLSTEQIILSL 308
      240 R-----ONGSSH-----NDLGSFPGDXEELTDEIADNITIGV 272
QY      309 LFAGHETSSVALAIFLQACPAVELEREHLEIARAKKEGSELNMDYRKMDPTQ 368
      273 IFAADTVASVSMILKYLAEKPNVLEAVTEBOMAI-RKDRKEGS-LTWGPTKKMPLTS 330
QY      369 CVINTELGNVYRFLRKALKDVRKGYDIPSGKYLPIVSAVHLNDSRYDQ 428
      331 RVIOETLVASLSTFEEAVEDEVESYLIPKGRVYVYTRRLNDYDRLPYD 390
QY      429 RMOQONNASSSGSFSTWGNMYMPFGGPRLCAGSELAKLEMAVFIHLVL 483
      391 RFE-----VAPKRTFMPFGNGTSCGNEELAKLEMSIMIHLLTTY 432

```

RESULT 8

```

T02739
probable cytochrome P450 At2g29090 [imported] - Arabidopsis thaliana
N:Alternate names: cytochrome P450 homolog T914.17
C:Species: Arabidopsis thaliana (mouse-ear cross)
C:Date: 24-Mar-1999 #sequence_revision 24-Mar-1999 #text_change 16-Feb-2001
C:Accession: T02739; D84692
R:Rounsley, S.D.; Lin, X.; Kaul, S.; Shea, T.P.; Fujii, C.Y.; Mason, T.M.; Shen, M.;
submitted to the EMBL Data Library, August 1998
A:Description: Arabidopsis thaliana chromosome II BAC T914 genomic sequence.
A:Reference number: Z14710
A:Accession: T02739
A:Status: translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 1-482 <ROU>
A:Cross-references: EMBL:AC005315; NID:g3461834; PID:g3461849
R:Lin, X.; Kaul, S.; Rounsley, S.D.; Shea, T.P.; Benito, M.I.; Town, C.D.; Fujii,
M.; Koo, H.; Moffat, K.S.; Cronin, L.A.; Shen, M.; VanKen, S.E.; Umeyam, T.; Tall
euss, D.; Merman, W.C.; White, O.; Eisen, J.A.; Salzberg, S.L.; Frazer, C.M.; Ven
Nature 402, 761-768, 1999
A:Title: Sequence and analysis of chromosome 2 of the plant Arabidopsis thaliana.
A:Reference number: AB4420; MUID:20083487; PMID:10617197
A:Accession: D84692
A:Molecule type: DNA
A:Status: preliminary
A:Residues: 1-482 <STO>
A:Cross-references: GB:AE002093; NID:g3461849; PIDN:ACG3235.1; GSPDB:GN00139
C:Genetics:
A:Gene: At2g29090; T914.17
A:Map position: 2
A:Introns: 80/2; 187/3; 238/3; 321/3; 351/3; 413/3; 453/2
C:Superfamily: Synechocystis cytochrome P450 slr0574; cytochrome P450 homolog
F:291-453/Domain: cytochrome P450 homolog <P45>

```

```

Query Match      22.8%; Score 611; DB 2; Length 482;
Best Local Similarity 29.8%; Pred. No. 4.5e-32;
Matches 153; Conservative 101; Mismatches 192; Indels 68; Gaps 13;
QY      16 PSLLSLFLILK-----RRNRKTRNLPKSGMPLGERTIGYIKPTATTLGDPMQ 69

```



```

Db      20 PALITLITVVVVVLLFKWMLHMKRQRLRPLPGSMGLPIGT---LRLYTEP--NSFFA 75
QY      70 QHVSXKGIYRSNLFGEPTIVSADAGLNRFITIQNERLRECSYPSIGITLGMWLVLY 129
Db      76 TRQNKRGDILFKHITLGCPCVMTSSPEAAMVLSKALHLPKPYPPSPKEMIGPEALFFHQ 135
QY      130 GDMHRDM-RSISLNFSLHARLRTILIKDVERHTLFVLDWMQONSIFSAQDEAKKTFENM 188
Db      136 GYHSTLRLKLVQSSFPMSALRPTV--SHIELIVQLTSSWTSOKSINTLEYKKRYAFDVA 193
QY      189 AKHISMDSGEEE---TEDLKREYVTFMKGVVSAPLNPFGAYHKAQSRATILKFERK 245
Db      194 ---HSAGDCKEPTTIDVILKLYORLEGVNSMDLDPGLTFHKSMAKRIELSELKRV 250
QY      246 MEERLIDKEDEEVEEYVTEDEAEKMSKDHVRKQRTDLDLGWL-----KSNLSTEQ 300
Db      251 IEKRRNGREE-----GGLGVLLGAKKQDRNLSSQ 283
QY      301 IIDLILSLFAGHETSSVAIALAIFLQACPKAVELEBEHLEI-ARAKKEGESELMND 359
Db      284 IADNTIGVIFATDTTASVLTMLKYLHDHPMLQEVSRQESINQIKKE--NRISME 341
QY      360 DYKKMDFTQCVINETLRGNVYRFLHKKALDVRKYKGDISGKRVLYTSAVHLNDSRY 419
Db      342 DTRKMLTTRVIGETLRASVLSFTFERAVODVEYDGLVLPKGMKVLPLFRIRHSSSEF 401
QY      420 DQPLFNFRWQOQONNGASSGSGSFTWGNMYFPGGPRLCAGSELAKEMAVFIHHL 479
Db      402 PDPEKFDPSREF-----VAPKPYTFMFGVSGVSCGSELAKEMLILLHL 448
QY      480 VLKFMWELAEDDOPFAF-PFYDFRNGPLPIRYSRI 512
Db      449 TTSFRWEYIGDEGIQIGPPVPPKGLPIRVTP 482

```

RESULT 9

```

C71417
cytochrome P450 d13695c - Arabidopsis thaliana
N:Contains: oxidoreductase (EC 1.-.-.-)
C:Species: Arabidopsis thaliana (mouse-ear cress)
A:Valley: columbia
C:Date: 10-Sep-1999 #sequence-revision 10-Sep-1999 #text-change 16-Jun-2000
C:Accession: C71417
R:Bevan, M.; Bancroft, I.; Bent, E.; Love, K.; Goodman, H.; Dean, C.; Bergkamp, R.; Dirk
avanagh, T.; Hempel, S.; Kotter, R.; Wellenreger, T.; Pohl, T.M.; Terry, N.; Giel
Nature 391, 485-488, 1998
A:Authors: Mueller-Auer, S.; Silvey, M.; James, R.; Montfort, A.; Pons, A.; Puigdemont
erhoft, A.; Moeres, T.; Jones, J.D.G.; Eneva, T.; Palme, K.; Benes, V.; Rechman, S.; And
C.; Chaitwalis, N.
A:Title: Analysis of 1.9 Mb of contiguous sequence from chromosome 4 of Arabidopsis thal
A:Reference number: A71400; MUID:98121113; PMID:9461215
A:Accession: C71417
A:Status: nucleic acid sequence not shown; translation not shown
A:Molecule type: DNA
A:Residues: 1-487 <BEV>
A:Cross-references: GB:Z97338; NID:g2244870; PIDN:CAB10309.1; PID:g2244888
C:Genetics:
A:Gene: d13695c
A:Map position: 4COP9-4G3845
C:Superfamily: Synchocystis cytochrome P450 slr0574; cytochrome P450 homology
C:Keywords: chromoprotein; heme; iron; metalloprotein; oxidoreductase
F:433/Binding site: heme iron (Cys) (axial ligand) #status predicted

```

```

Query Match      21.7%; Score 581; DB 1; Length 487;
Best Local Similarity 27.3%; Pred. No. 4e-30;
Matches 143; Conservative 100; Mismatches 205; Indels 76; Gaps 11;

```

```

QY      17 SLTSLILFLILKRRRRTKRFNLPKSGMPLGRTIGLKRYATATLGDPMQOHVSKYG 76
Db      3 SLFLVILFHWYQMKRPKTKNGKLPESMGEPFIFGEFFKPHDALQFSTFKDRLVRF 62
QY      77 -----KIYRSNLFGEPTIVSADAGLNRFITIQNERLRECSYPSIGI-----LG 121

```

```

Db      63 ADFSSIHLSFEPTSLFGDKALISMDELN-----LEKANSIPGVTKSVIRLFG 112
QY      122 KSMVLVVDMDHRDRSRISLNLHARLRTILIKDVERHTLFVLDWMQONSIFSAQDEAK 181
Db      113 ENNLPLQSKESHKRNRLTFOLLPOGLKSRNIEVDVLAARYMEGARNGVLDVKTSS 172
QY      182 KFTFMMAKHIM-SMDPGEETEOCLKREYVFMKGVVSAPLNPFGAYHKAQSRATILKFERK 232
Db      173 KILIGCLAKKTVGEMEP--EAKELALWRFQSGWFFPFLNLPETGVYKMKVLYFOYT 230
QY      223 -----QSRATILKFERKMEERKIDKEDEEVEEYVTEDEAEKMSKDHVRKQRTDLD 286
Db      231 EADISWQARKRMKILRKTIVLTKRASGEELGEFNIIRIGEGE----- 274
QY      287 LGWVLKHSNLSSTEQIIDLILSLFAGHETSSVAIALAIFLQACPKAVELEBEHLEIAR 346
Db      275 -----GETMVENAWEITYTFELVANETTPRIILAAVVKFISDHPKQOELORHEETVR 328
QY      347 AKKELGESELMNDYKKMDFTQCVINETLRGNVYRFLHKKALDVRKYKGDIPSGMKVL 406
Db      329 GKAE-KEGGLTWEDYKSMHFTQVINSRLRITSTAPVYKLVLEHDPQGYTTPAGWTFM 387
QY      407 PVISAVHLNDSRYDQPLFNFRWQOQONNGASSGSGSFTWGNMYFPGGPRLCAGSE 466
Db      388 G-YPHIHNSKEDYDPAFNPFRWEGKDLGAIYS-----KTFIPFGAGRILCVGAE 437
QY      467 LAKLEMAVFIHHLVLKFMWELAEDDOPFAF-PFYDFRNGPLPIRYSRI 510
Db      438 FAKQMAVFIHHL-FRYRMSKSGTTIIRSPMLMFGPCDVOIS 480

```

RESULT 10

```

A84859
probable cytochrome P450 [imported] - Arabidopsis thaliana
C:Species: Arabidopsis thaliana (mouse-ear cress)
C:Date: 02-Feb-2001 #sequence-revision 02-Feb-2001 #text-change 23-Mar-2001
C:Accession: A84859
R:Lin, X.; Kail, S.; Rounsley, S.D.; Shea, T.P.; Benito, M.I.; Town, C.D.; Fujii, G.Y
M.; Koo, H.; Moffat, K.S.; Cronin, L.A.; Shen, M.; VanAken, S.E.; Umayam, L.; Tallo
euss, D.; Nierman, W.C.; White, O.; Eisen, J.A.; Salzberg, S.L.; Fraser, C.M.; Venter
Nature 402, 761-768, 1999
A:Title: Sequence and analysis of chromosome 2 of the plant Arabidopsis thaliana.
A:Reference number: A84420; MUID:20083487; PMID:10617197
A:Accession: A84859
A:Status: Preliminary
A:Molecule type: DNA
A:Residues: 1-485 <STO>
A:Cross-references: GB:AE002093; NID:g4512670; PIDN:AAD21724.1; GSPDB:GN00139
C:Genetics:
A:Gene: At2g42850
A:Map position: 2
C:Superfamily: Synchocystis cytochrome P450 slr0574; cytochrome P450 homology
C:Keywords: heme; iron; metalloprotein
F:432/Binding site: heme iron (Cys) (axial ligand) #status predicted

```

```

Query Match      21.6%; Score 579.5; DB 2; Length 485;
Best Local Similarity 28.8%; Pred. No. 5e-30;
Matches 147; Conservative 107; Mismatches 203; Indels 53; Gaps 11;

```

```

QY      12 LILPLSLILFL-----ILKRRNRKTRFNLDPKSGMPLGRTIGLKRYAT--TL 64
Db      14 LCIATFISTLFFPKKHKHRTTKIQQKK--LIPGEGDLWIGETWDFYAKQSNRYF 71
QY      65 GPFMOHVSQKTIYRSNLFGEPTIVSADAGLNRFITIQNERLRECSYPSIGITIGKWS 124
Db      72 EDFVNPRIKIKGNIKRTIMGSPITIVGAEANRLILNERSLVVSSMPSSVQDLMGNC 131
QY      125 MLVVGDMHDKRSISLNLHARLRTILIKDVERHTLFVLDWMQONSIFSAQDEAKFT 184
Db      132 IMAGGEKHRLVLRGIVANSLSYIGLSLIPKLCDTVFKHETEMKKEEISLYLSAKVLT 191
QY      185 FNLMAKHI--MSMDPGEETEOCLKREYVFMKGVVSAPLNPFGAYHKAQSRATILKFI 242

```

```
Db 192 FTVVECLYGIKVELGMLEV-----FEVLEGVFALVEPPCSFAAKAKARLEIEFVL 245
Qy 243 ERKMEERKLIDKEEDOEVEEVKTEDEAMSKSDHAKORTDDLLGWALKHSNLSLTOIL 302
Db 246 VGKVRKREKREMEKGAEE-----KPNITLFSRLVEBELKGV-----ITEEVY 287
Qy 303 DLILSLFAGHETSSVAIALAIFLQACPKAVEELREHELEIARAKKEIGSE-LNMDDY 361
Db 288 DNMVLVFAAHDITSYAMSMTEFKMLAQHPTCRDITLQEHAOI---KANKGEYELVEDV 344
Qy 362 KKMDFOCVINEITRLGNVRFELHRAKLDVRYKGYDIPSGMKVLPVLSAVHLDSRDO 421
Db 345 KKMYSYQVAVETIRLSPPIGSRKAAVADIDYGYITPKMKILMTTYGTHNPEIFOD 404
Qy 422 PNLNPMRMOQONNGASSSGSSTWGNMYPFGGPRLCAGSELAKLEMAVFIHLVL 481
Db 405 PMSDFPTRFDPKPIQAYT-----YLPFGGGRRLCAGHQLAMLSILVEHFHYVT 451
Qy 482 KFNELALDDQPRFAFPVDFPN-GLPIRVS 510
Db 452 GFDMSLVYPPDETISMDFLPFSLQMPKIS 481
```

RESULT 11

hypothetical protein [Imported] - Arabidopsis thaliana
C:Species: Arabidopsis thaliana (mouse-ear cress)
C>Date: 02-Mar-2001 #sequence_revision 02-Mar-2001 #text_change 31-Mar-2001
C:Accession: H86185
R:Theologos, A.; Ecker, J.R.; Palm, C.J.; Federpiel, N.A.; Kaul, S.; White, O.; Alonso,
Chin, C.W.; Chung, M.K.; Conn, L.; Conway, A.B.; Conway, A.R.; Creasy, T.H.; Dewar, K.;
ansen, N.F.; Hughes, B.; Huizar, L.
Nature 408, 816-820, 2000
C.A.: Li, J.H.; Li, Y.; Lin, X.; Liu, S.X.; Liu, Z.A.; Luos, J.S.; Maiti, R.; Marziani,
Rizzo, M.; Rooney, T.; Rowley, D.; Sakano, H.
A:Authors: Salzberg, S.L.; Schwartz, J.R.; Shinn, P.; Southwick, A.M.; Sun, H.; Tallon,
ker, M.; Wu, D.; Yu, G.; Fraser, C.M.; Venter, J.C.; Davis, R.W.
A:Title: Sequence and analysis of chromosome 1 of the plant Arabidopsis.
A:Reference number: A86141; MUID:21016719; PMID:11130712
A:Accession: H86185
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-490 <STO>
A:Cross-references: GB:AE005172; NID:92388581; PIDN:AB71462.1; GSPDB:GN00141
C:Genetics:
A:Map position: 1
C:Superfamily: Synchocystis cytochrome P450 slr0574; cytochrome P450 homology

```
Query Match 21.6%; Score 579.5; DB 2; Length 490;  
Best Local Similarity 28.6%; Pred. No. 5e-30;  
Matches 133; Conservative 107; Mismatches 192; Indels 33; Gaps 9;  
Qy 34 KTRFNLPKSGMPLFETIGYLKPYATTLGDMQOQVSKYK--IYRSNLFGEPTIYS 91  
Db 40 ENRHILPFGDLGMPFIENMLSLFAKTSIDPDSTRTLLIKRGKGIYKAMFGPSTIV 99  
Qy 92 ADAGLNRLIQLNEGRLEFCSYRISGILGKWSMLVVGDMHMRMSISLNFSLHARLT 151  
Db 100 TTSOTCARVLTDOD-AFKPWPSTWMLIGKRSFVGSFEHKKRLRLTARVAGHEALS 158  
Qy 152 ILKDVARRHLFVLDWMQONSIFSAODEAKKFTNLAQKIMSDPGEETDQKKEYVT 211  
Db 159 TYPIREENVITVLDKTKGEEFFLTHLRKLFRII-MYFLSSESNVADALEREYTA 217  
Qy 212 FMGVASAPLNPCTAVALKALOSRATILKFIERKMERKLDIKREDOEEVYKDEDEM 271  
Db 218 LNTGVRAAVNIPEFAHRAKAKKTLVAAPOSTVTER-----NORKNIISSKKDML 271  
Qy 272 SKSDVAKORTDLDLGLWVAKHSNLSLTOILDLISLFGHETSSVAIALAIFLQACP 331  
Db 272 DNLNLNKK-----GKTLDEDEILIDVILMTLNMGHSSGHTIMATYVFLQENP 320
```

```
Qy 332 KAVEELREHELEIARAKKEIGSELMNDYKKMDFQCYINTELRLGNVRFELHRAKLD 391  
Db 321 EYLQAKARQEMILKSRPE-GQKGLSLKTEFKMEFLISQVYDETLAVITTSFLAFREAKTD 379  
Qy 392 VYKGYDITSGKAVLPVISAVALHDSNRITQPNLFNWRMOQONNGASSSGSFTWGN 451  
Db 380 VEMNGLLIPKGRKVLWFRDVIHIDPEVFPDPKRFDPARM---DNG-----FVPRAGA 428  
Qy 452 YMPFGGPRLCAGSELAKLEMAVFIHLVLKFNWELAEEDQPRFA 496  
Db 429 FLPRGASHLCPGNDLAKLEISFLIHLFLTKQVAKSNBECPRMY 473
```

RESULT 12

cytochrome P450-like protein - Arabidopsis thaliana
N:Alternate names: protein F14D17.40
C:Species: Arabidopsis thaliana (mouse-ear cress)
C>Date: 02-Jun-2000 #sequence_revision 02-Jun-2000 #text_change 18-Aug-2000
C:Accession: T48973
R:Jordan, N.; Bangert, S.; Wiedelmann, R.; Voss, H.; Unseld, M.; Mewes, H.W.; Ruedd,
A:Reference number: 225008
A:Accession: T48973
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-455 <JOR>
A:Cross-references: EMBL:AL353992; GSPDB:GN00061; ATSP:F14D17.40
A:Experimental source: cultured Columbia; BAC clone F14D17
C:Genetics:
A:Gene: ATSP:F14D17.40
A:Map position: 3
A:Introns: 70/2; 178/3; 229/3; 312/3; 343/3; 346/1; 381/3
C:Superfamily: Synchocystis cytochrome P450 slr0574; cytochrome P450 homology

```
Query Match 21.5%; Score 577; DB 2; Length 455;  
Best Local Similarity 28.9%; Pred. No. 6.6e-30;  
Matches 145; Conservative 84; Mismatches 173; Indels 100; Gaps 10;  
Qy 32 NRKTRFNLPKSGMPLFETIGYLKPYATTLGDMQOQVSKYKITSNLFGEPTIYS 91  
Db 28 NPKSGKLRPGSGMPLFETIGYLFDFRPIGYEISPLKMKMLRYGLPFTNITIGVTVS 87  
Qy 92 ADAGLNRLIQLNEGRLEFCSYRISGILGKWSMLVVGDMHMRMSISLNFSLHARLT 151  
Db 88 TDKVNMETLROENKSFILSYDGLMKPLGKDSLFLKIGNIKHIKQITLHLLSSGLKR 147  
Qy 152 ILKDVARRHLFVLDWMQONSIFSAODEAKKFTNLAQKIMSDPGEETDQKKEYVT 211  
Db 148 KILKMDRYTRREHLSKAKTGRLDVDAVSKLIIAHLPRKMS-----NLKPOTQA 198  
Qy 212 FMKGV-----VSAPLNPCTAVALKALOSRATILK--FIERKMERKLDIKE 255  
Db 199 KIMGIFKAFPTFMRFTSYLSIGKGLYNLM--ACREGKREIKDITMKTSEK----- 252  
Qy 256 EDQEEVEVKTDEDEAMSKSDHAKORTDDLLGLWALKHSN-----LSTEQILDLISLLE 310  
Db 253 -----DDFLNTAIEESEKAGELLNENAIITLLPTLSC 284  
Qy 311 AGHETSSVAIALAIFLQACPKAVEELREHELEIARAKKEIGSELMNDYK-KMDFQOC 369  
Db 285 VTQDTTSKALICLAVFLLENPVVLELKEH-EVILSEKDEEGGTWEYHKKMFTN- 342  
Qy 370 VINETLRLGNVRFELHRAKLDVRYKGYDIPSGMKVLPVLSAVHLDSNRITQPNLFNWR 429  
Db 343 -----MKSGTIPRAGVIMYIIPSYVNHDPRETYENPFEENPWR 379  
Qy 430 WQOONNGASSSGSSTWGNMYPFGGPRLCAGSELAKLEMAVFIHLVLKFNWELAE 489  
Db 380 WEGKELRAGS-----KTFVFGGLRGQACAGAFARLQISVFLHHLVTVNPSLHQ 429  
Qy 490 DDQPRAPFPVDFPNGLPIRVS 511
```

Db 430 DCEVLKVPAAHLPGNISINISK 451

RESULT 13

B84733

probable cytochrome P450 (imported) - Arabidopsis thaliana

C:Species: Arabidopsis thaliana (mouse-ear cress)

C:Date: 02-Feb-2001 #sequence_revision 02-Feb-2001 #text_change 16-Feb-2001

C:Accession: B84733

R:Lin, X.; Kaul, S.; Rounsley, S.D.; Shea, T.P.; Benito, M.I.; Town, C.D.; Fujii, C.Y.; M.; Koo, H.; Moffat, K.S.; Cronin, L.A.; Shen, M.; Vanaken, S.E.; Umayan, L.; Tallon, L.; Nuss, D.; Nierman, W.C.; White, O.; Eisen, J.A.; Salzberg, S.L.; Fraser, C.M.; Venter, J.

A:Title: Sequence and analysis of chromosome 2 of the plant Arabidopsis thaliana.

A:Reference number: A84420; MUID:20083487; PMID:10617197

A:Accession: B84733

A:Status: preliminary

A:Molecule type: DNA

A:Residues: 1-489 <STO>

A:Cross-references: GB:AE002093; NID:93831452; PIDN:AAC69934.1; GSPDB:GN00139

A:Gene: At2g32440

A:Map position: 2

C:Superfamily: Synecocystis cytochrome P450

Query Match

Best Local Similarity 20.4%; Score 547.5; DB 2; Length 489;

Matches 146; Conservative 105; Mismatches 211; Indels 67; Gaps 14;

Db 8 TLPLLLPLLSLLFLI--LKKRR-----KTRNLPKSGSGWFLGTTGYLK 57

4 TGLILMFP-LIILGLFVLKWLKRVNWIYVSKLEKKHYLPKGLGMPVIGNMSPFLR 62

Qy 58 PYRATITGDFMOQVHVSRYK--LYRSNLEGEPTIVSADAGLNRFILQNGRLEFCSPYS 115

Db 63 AFTSDPESFIOSTYIRGRTGYIKAHMGYPCVLTTPETCRVLTDD--AHLGWPS 121

Qy 116 IGGILGKMSLVYVGMHMDRMSISLNFSLHARLITLLKDYERHTLFVLDWQONSIRS 175

Db 122 TMLIGKRSVGVISFEHRLRLTSAPVNGPGLSVYIOFIEETVNTDLKMSKMGTE 181

Qy 176 AODEAKFTFNLMKAIMSDPEETEDLKEKYVFMKGVSAFLNPGTAYHKLQSR 235

Db 182 FLHLKRLTFKYI--MTFLSSSEHMDLEREYTNLNGVRAKGINLPGFYHRLAKR 240

Qy 236 ----ATLKFIERKMEERKLDI--KEEQEEVEKTEDEAEKSKSDHVKQRTDDLLGW 289

Db 241 KKLVAAPFOSIVTNRNORONISNNKMDLMDLIDVDE-----NGRYLDD----- 286

Qy 290 VLKHSNLSLEQILDLISLFFAGHETSVAIALALFFLOACPKAVEELREHLEIARAK 349

Db 287 -----EETIDLLIMYANGHSSGHLTMWATILMOEHMILQKKEQERL--VKK 335

Qy 350 ELGESELMNDYKKKDDFTQCVINETLRGNNVRLHRRKALDVRYKGDIPSGKVLPIV 409

Db 336 RAPGCKTLTKETRENVYLSOYDELRLVTFSLAFRAKSDYQMDGYIIIRKQKYLTFW 395

Qy 410 SAVHLDNSRYDQPNLFNPMRQOONNGASSSGSFSFGWNNYMPFGGPPKLCAGSELAK 469

Db 396 RNVLHDPETLPDKKFDSSKME-----GYPKAGTFLPFGISHLGCGNDLAK 443

Qy 470 LEMAVFHLYLVKFWELAEEDDOPAPFVDFPGLP-----INVRIL 513

Db 444 LEISIFLHFLKLYKVERSNPGCPVMF-----LPNNRKNDCNLAITITRM 488

RESULT 14

D96813

hypothetical protein T30F21.17 (imported) - Arabidopsis thaliana

C:Species: Arabidopsis thaliana (mouse-ear cress)

C:Date: 02-Mar-2001 #sequence_revision 02-Mar-2001 #text_change 19-Apr-2002

C:Accession: D96813

R:Theologis, A.; Ecker, J.R.; Palm, C.J.; Federspiel, N.A.; Kaul, S.; White, O.; Alonso, C.W.; Chung, M.K.; Conn, L.; Conway, A.B.; Conway, A.R.; Creasy, T.H.; Dewar, ansen, N.F.; Hughes, B.; Huizart, L.

A:Authors: Hunter, J.L.; Jenkins, J.; Johnson-Hopson, C.; Khan, S.; Khaykin, E.; Kim, C.A.; Li, J.H.; Li, Y.; Lin, X.; Liu, S.X.; Liu, Z.A.; Luoro, J.S.; Malt, R.; Marjia, Rizzo, M.; Rooney, T.; Rowley, D.; Sakano, H.

A:Autors: Salzberg, S.L.; Schwartz, J.R.; Shim, P.; Southwick, A.M.; Sun, H.; Tallon, ker, M.; Wu, D.; Yu, G.; Fraser, C.M.; Venter, J.C.; Davis, R.W.

A:Title: Sequence and analysis of chromosome 1 of the plant Arabidopsis.

A:Reference number: A86141; MUID:21016719; PMID:11130712

A:Accession: D96813

A:Status: preliminary

A:Molecule type: DNA

A:Residues: 1-460 <STO>

A:Cross-references: GB:AE005173; NID:94836883; PIDN:AAD30586.1; GSPDB:GN00141

A:Gene: T30F21.17

A:Map position: 1

C:Superfamily: Synecocystis cytochrome P450

Query Match

Best Local Similarity 18.6%; Score 497.5; DB 2; Length 460;

Matches 133; Conservative 102; Mismatches 205; Indels 67; Gaps 12;

Db 13 LLPLLSLLPLLLILKRRNRKTRFNLPKSGSGWFLGTTGYLKPYRATITGDMQOHV 72

9 MLNVALVVRISHLWLRMSNPRCPKGLPFGSGWFLGTTGYLKPYRATITGDMQOHV 68

Qy 73 SKYGYKRSNLEGEPTIVSADAGLNRFILQNGRLEFCSPYSISIGLIGKMSLVYVDM 132

Db 69 IRGPLEFTNIGSTVSTDPDVTHOIFROHNTSEFELGYPDIEFVKVGNLFLKEVEFI 128

Qy 133 HRDMSISLNFSLHARLITLLKDYERHTLFVLDWQONSIFSADDEKAKTFLNMAKH 192

Db 129 HYYLQITMQLIGSGGLQTMGNNDKATROHINSIASOGSFNVRKEVENLVAYMPKL 188

Qy 193 MSMDGEETEDLKEKYVFMKGVSAFLNPGTAYHKLQSRATILFKIERKMEERKLD 252

Db 189 ISNLRPEQSKLIDN-----LNAFNIDMFKSF--LRISTKAVYKALKSR--- 231

Qy 253 IKEDQEEVEKTEDEAEKSKSDHVKQRTDDLLGWLKH-----SNLSTQILDLIS 307

Db 232 -----BEALQVAKDVLMAKKEIREKO---EDFLNLLLELEKDGSGFFDQSAINLFL 281

Qy 308 LLFPGHETSVAIALALFFLOACPKAVEELREHLEIARAKKEISELMNDYK--KMD 366

Db 282 LAFALREGTSCALAVKFSKDPKVLAEKREKALVDNRKD--KEAGVSWERTRMNTF 340

Qy 367 TOCVINELIRGNNVRLHRRKALDKVYK--YDIPSGWVLPVISAVALDNSRYDQPNLF 425

Db 341 TNMVSNEVLRLANTPLLPFRKAVDVEIKRY-----LLEVIHMGFG----- 382

Qy 426 NPMRQOONNGASSSGSFSFGWNNYMPFGGPPKLCAGSELAKLEMAFVHLYLVKRN 484

Db 383 --WLMQ-----GKEMINGSKTFMAFGVNLCAEERSLQMAFLHNLVAYYD 429

Qy 485 WELAEEDDOPAPFVDFPGLP-----INVRIL 511

Db 430 FSNVQDSERTIRSPHOTKDLINTISO 456

RESULT 15

T02263

cytochrome P450 DWARF3 - maize

N:Contains: oxidoreductase (EC 1.-.-.-)

C:Species: Zea mays (maize)

C:Date: 10-Sep-1999 #sequence_revision 10-Sep-1999 #text_change 21-Jul-2000

C:Accession: T02263

R:Winkler, R.G.; Helentjaris, T.

A:Title: The maize dwarf3 gene encodes a cytochrome P450-mediated early step in gibber

A;Accession: T02263
A>Status: preliminary

A/Status: preliminary; translated from GB/EMBL/DBJ
A/Molecule type: mRNA

A: Molecule type: mRNA

A;Residues: 1-519 <WIN>

A; Cross-references: EMBL:U32579; NID:g987266; PIDN:AAC49067.1; PID:g987267

A; Experimental source: strain B73

A, Experimental

A;Gene: dwarf3

A;Gene: *dwa*
C;Function:

A:Description: Involved in an early step in gibberellin biosynthesis

A;Description: Involved in an early A;Pathway: gibberellin biosynthesis

A;pathway: gibberellin biosynthesis
C:Superfamily: Synechocystis cytochrome P450 slr0574: cytochrome P450 homolog

C/Superfamily: Cytochrome P450 slr0574; cytochrome P450 homology
C/Keywords: oxidoreductase

C;Keywords: oxidoreductase
E:325-488/Domain: cytochrome P450 hemojcnr <P45>

Query Match 18.38; Score 491.5; DB 1; Length 519;

Best Local Similarity 28.2%; Pred. No. 2.7e-24;

Local similarity 20.28; Freq. NO. 2.7E-24;
 Matches 129; Conservative 91; Mismatches 189; Indels 49; Gaps 11;

36 RNLPRGKSGWPLGETIGYLKPYTATLTGDFMQQHVSKYQG - IYRSNLFGEPTI-VSA 92

Db 69 KARLPCEMGWPLVGGMWAFRAFKSGKPDAFIASFVRREGRTGVYRSEFMSSPTVLVTT 128

QY 93 DAGLNRFI LONEGRLEFECSPRISIGGILGKWSMLVLVGDHHRDMRSISLNFLSHARLTI 1522

Db 129 AEGCKOVLMDDA--EVTGMPKATVALVGRSEFVAMPYDEHRRIRKLTAPINGEDALTG 186

QY 153 LKDVERTLEVLDSW-OONSIFSADPEAKKFTENIMAKHIMSDPGEFEETEOI.KKEYVT 211

Db 187 YLPFIDRTVTSRLRAWADHGSGVEFTELFRMTFKIVO-TFIGGADQATTRALERSYTE 245

212 FMKGVSAPI.NPGTAYHKAIOSRATTI.KTEFRKMFEFRIDKE-----FDOEEFEVKT 365

Db 246 TNYGMRAMATNPEGAVRGALPABPRTYAVTLOGVI DEBPABAKVCSCGVMMDBI TEA 305

00 266 EDEAFMSKSDHVBKOPDNDI I CWTI KHSNI STEOIT DI TI CI I PACHEMGCWATAT ATN 33E

```

:| | | | |
306 DDE-----BCBUI DDD-----ETDUI DDDI MACHGCCGUTWWTWWT 343
DB

```

336 ET AL.

27	28	29	30	31	32	33	34	35	36	37	38	39	40	41	42	43	44	45	46	47	48	49	50	51	52	53	54	55	56	57	58	59	60	61	62	63	64	65	66	67	68	69	70	71	72	73	74	75	76	77	78	79	80	81	82	83	84	85	86	87	88	89	90	91	92	93	94	95	96	97	98	99	100																										
1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26	27	28	29	30	31	32	33	34	35	36	37	38	39	40	41	42	43	44	45	46	47	48	49	50	51	52	53	54	55	56	57	58	59	60	61	62	63	64	65	66	67	68	69	70	71	72	73	74	75	76	77	78	79	80	81	82	83	84	85	86	87	88	89	90	91	92	93	94	95	96	97	98	99	100

00C FBI MEMPHIS TEL AUGUST TWO FOUR LAST AND BUREAU
00C FBI WASHINGTON TEL AUGUST TWO FOUR LAST.

200 KIMBANDY KIMNOLOLE COMNAV VLSRV HEDNCKLDFNLEINFWKWQZQZNNQAS353535357 443

402 AGAARDVFVNGILILFNGMAVQLMIKRSVHMDPVVIFDPVKFDPKNE-----GHSRKAGTF 456

440 SIMGNINMPFGGPRCAGSELAKLEMAVFIHHLVKE 483
: | | | | : : | | | : | | : :
.

Search completed: June 10, 2003, 21:59:31
Job time : 30 secs

Job time : 30 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: June 10, 2003, 21:51:48 ; Search time 18 Seconds

(without alignments)
1182.075 Million cell updates/sec

Title: US-09-502-426B-2

Perfect score: 2681
Sequence: 1 MFETEHHTLPLILPLSLLS.....FAFPYDFPGLPIRVSRL 513

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 112892 seqs, 41476328 residues

Total number of hits satisfying chosen parameters: 112892

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%

Maximum Match 100%
Listing first 45 summaries

Database: SwissProt_40.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	978.5	36.5	472	C901_ARATH	042569 arabidopsis
2	864	32.2	524	C90C_ARATH	09m066 arabidopsis
3	761.5	28.4	464	CP85_LYCES	043147 lycopersico
4	579.5	21.6	490	C881_ARATH	023051 arabidopsis
5	503	18.8	492	CP26_BRARE	P79739 brachydanio
6	491.5	18.3	519	C881_WAIZE	043246 zea mays (m
7	479.5	17.9	444	C120_SYNY3	059990 synecocyst
8	461	17.2	497	CP26_MOUSE	055127 mus musculu
9	446	16.6	512	C26A_HUMAN	09m163 homo sapien
10	444.5	16.6	492	CP26_XENLA	093323 xenopus lae
11	439.5	16.4	492	CP26_CHICK	09m163 homo sapien
12	438	16.3	497	CP26_HUMAN	043174 homo sapien
13	324.5	12.1	520	YRV2_CAEEL	027514 caenorhabdi
14	321.5	12.0	518	YRVA_CAEEL	027510 caenorhabdi
15	316	11.8	500	C912_ARATH	065790 arabidopsis
16	312.5	11.7	504	CP32_RAT	P05183 rattus norv
17	312	11.6	501	CP35_MOUSE	054749 mus musculu
18	310.5	11.6	520	YRV1_CAEEL	027513 caenorhabdi
19	303	11.3	501	CP76_MOUSE	054750 mus musculu
20	302	11.3	494	CPA8_MESAU	P24454 mesocricetu
21	299.5	11.2	492	CP8C_RAT	P33772 rattus norv
22	297	11.1	491	CPB9_MOUSE	P12790 mus musculu
23	294.5	11.0	491	CPB4_RABIT	P00178 oryctolagus
24	294	11.0	506	C892_ARATH	042602 arabidopsis
25	293	10.9	491	CPB1_RAT	P00176 rattus norv
26	291.5	10.9	530	CP51_YEAST	P10614 saccharomyc
27	290.5	10.8	502	C83A_ARATH	P48421 arabidopsis
28	289	10.8	503	CP39_RAT	P51538 rattus norv
29	288.5	10.8	492	CPB1_MOUSE	O55071 mus musculu
30	288.5	10.8	503	CP33_HUMAN	P05184 homo sapien
31	288.5	10.8	538	CP18_DROME	O95078 drosophila
32	288	10.7	459	CPYN_ANASP	P29980 anabaena sp
33	287.5	10.7	520	YRV5_CAEEL	027517 caenorhabdi

34	286.5	10.7	533	CP51_CANGA	P50859 candida gla
35	285.5	10.6	502	CPJ2_HUMAN	P51589 homo sapien
36	284.5	10.6	505	C762_SOLME	P37122 solanum mel
37	282	10.5	514	CP41_VICSA	O81117 vicia sativ
38	281.5	10.5	499	CP81_ARATH	O55782 arabidopsis
39	281.5	10.5	504	CP3G_MOUSE	O64481 mus musculu
40	281	10.5	528	CP51_CANTR	P14263 candida tro
41	280.5	10.5	504	CP3B_MOUSE	O64459 mus musculu
42	280	10.4	499	C771_SOLME	P37123 solanum mel
43	280	10.4	500	CPJ1_RABIT	P52786 oryctolagus
44	279.5	10.4	513	C773_SOYBN	O48928 glycine max
45	279.5	10.4	519	YRV8_CAEEL	Q27520 caenorhabdi

ALIGNMENTS

RESULT 1
ID C901_ARATH STANDARD; PRT; 472 AA.
AC Q42569;
DT 15-DEC-1998 (Rel. 37, Created)
DT 15-DEC-1998 (Rel. 37, Last sequence update)
DE 16-OCT-2001 (Rel. 40, Last annotation update)
DE Cytochrome P450 90A1 (EC 1.14.-.-).
GN CYP90A1 OR CYP90 OR CPD OR AT5G05690 OR MJJ3.9.
OS Arabidopsis thaliana (Mouse-ear cress).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;
OC eurosids II; Brassicales; Brassicaceae; Arabidopsis.
OX NCBI_TaxID:3702.
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=cv. Columbia;
RX MEDLINE=96200769; PubMed=8612270;
RA Szekeres M., Nemeth K., Koncz-Kalman Z., Mathur J., Kauschmann A.,
Altman T., Redei G.F., Nagy F., Schell J., Koncz C.,
RT "Brassinosteroids rescue the deficiency of CYP90, a cytochrome P450,
RT controlling cell elongation and de-etiolation in Arabidopsis.";
RL Cell 85:171-182(1996).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=cv. Columbia;
RX MEDLINE=97471969; PubMed=9330910;
RA Sato S., Kotani H., Nakamura Y., Kaneko T., Asanizu E., Fukami M.,
MYAJIMA N., Tabei S.,
RT "Structural analysis of Arabidopsis thaliana chromosome 5. I. Sequence
RT features of the 1.6 Mb regions covered by twenty physically assigned
RT pl clones.";
RL DNA Res. 4:215-230(1997).
CC 1- SIMILARITY: BELONGS TO THE CYTOCHROME P450 FAMILY.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL collaboration
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement. (See <http://www.isb-sdb.ch/announce/>
CC or send an email to license@sdb.ch).DR EMBL: X87367; CAA60793.1; -
DR EMBL: X87368; CAA60794.1; -
DR EMBL: AB005237; BAB09663.1; -
DR InterPro: IPR001128; Cytochrome_P450.
DR Pfam: PF00067; p450.1.
DR PROSITE: PS00086; CYTOCHROME_P450; 1.
KW Oxidoreductase; Monooxygenase; Transmembrane; Heme; Multigene family.
FT TRANSMEM 1 21
FT BINDING 418 418 HEME (BY SIMILARITY).
FT POTENTIAL.
SQ SEQUENCE 472 AA; 53785 MW; 41A73F46D64E343F CRC64;
Query Match 36.5%; Score 978.5; DB 1; Length 472;
Best Local Similarity 41.9%; Pred. No. 1.7e-51;

Matches 210: Conservative 79: Mismatches 171: Indels 41: Gaps 8:

QY 12 LLLPSSLILFLILKRRNRKRNLPKSGMPFJGTYGKPYATTATLGDGFMDOH 71
 Db 7 LLLSSISAGFL---LLLRTRYRRMGGLPPGSLGPIPLGEGPOLIGAVKTEPEPFI 63
 QY 72 VSKYKIRSNMFEPTIVSADAGNRFPLONEGLFCGYSRISGILGKRSMLVGD 131
 Db 64 VARGSVMTLFGPEPTIFSDPEPNRVLONEGLFCGYSRISGILGKRSMLVGD 123
 QY 132 MHRDMRSISLNLFLSHARLRTLLKDNVERHTFLVDSMOONSIFSAODEAKKFTFNLMAKH 191
 Db 124 LKRNHSLTMSFANSIIKDLMLDIDLVRNLDMSRVL---MEKAKTTFEITYKQ 181
 QY 192 IMSNDPGEETDOLKEVYTMKGVVAPLPGYAIKALOSRATILFKRKKEERKL 251
 Db 182 LMSDPG-EMSESLRKEFLVIEGFSPLPLFFSTYKKAIOAR-----RKVAEALT 232
 QY 252 DKEEDDEEVEVTKDEAEMSKSDHVRKQRTDDLLGVNLKHSNLTSTQIIDLISLFA 311
 Db 233 VVWAKRREEEGAE-----RKMDLAALLAADGFSDEIDELVALLVA 278
 QY 312 GHETSSVAIALAIFFLACPAVELREHLEIRAKKELESEELNMDYKMDTCVY 371
 Db 279 GYETSTIMTLAVKFLFTPLALQKHEHKIRAKMD--SYSEMDDYKSMPTOCVY 336
 QY 372 NETIARLGNVAFLLRKALKDVRIGYDIPSCMKVLPVISAHLDSRYDQPLFNPWKQ 431
 Db 337 NETIRVANIIGGVFRRAATDVEIKGYIPKMGVSSPRAVHLDPNHFEDARTFNPWKQ 396
 QY 432 QONNGASSGSGSSTGMGN--YMPGSGPRCAGSELAKLMAYIHLVLKFNELAD 490
 Db 397 -----SNSTTGPSSNVFTFGGSPRLCPGIELARVALSVFLHRLVTFGFSVAPQ 446
 QY 491 DQFAFPFVDFPNCPIRVS 511
 Db 447 DKLVEFPTRTQKRYPIFKR 467

RESULT 2
 C90C_ARATH STANDARD: PRT: 524 AA.

AC Q9M066: 023242;
 DT 16-OCT-2001 (Rel. 40, Last sequence update)
 DT 16-OCT-2001 (Rel. 40, Last sequence update)
 DT 15-JUN-2002 (Rel. 41, Last annotation update)
 DE Cytochrome P450 90C1 (EC 1.14.-.-) (ROTUNDIFOLIA3).
 GN ROT3 OR CYP90C1 OR AT4G36380 OR C7A10.980 OR F23E13.220.
 OS Arabidopsis thaliana (Mouse-ear cress).
 OC Eukaryota: Viridiplantae: Streptophyta: Embryophyta: Tracheophyta;
 OC Spermatophyta: Magnoliophyta: eudicotyledons: core eudicots: Rosidae;
 OC Eucosids II: Brassicales; Brassicaceae; Arabidopsis.
 OC NCBI_TaxID=3702;
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=98361880; Pubmed=9694802;
 RA Kim G.-T., Tsukaya H., Uchimiya H.,
 RT "The ROTUNDIFOLIA3 gene of Arabidopsis thaliana encodes a new member
 of the cytochrome P-450 family that is required for the regulated
 polar elongation of leaf cells.";
 RL Genes Dev. 12:2381-2391(1998).
 RN [2]
 RP SEQUENCE FROM N.A.
 RX STRAIN=Cv. Columbia;
 MEDLINE=20083488; Pubmed=10617198;
 RA Mayer K.F.X., Schaefer C., Wandutt R., Murphy G., Volckaert G.,
 RA Pohl T., Duesterhoeft A., Stiekema W., Entian K.-D., Terryn N.,
 RA Harris B., Ansong W., Brandt P., Grivell L.A., Rieger M.,
 RA Weichselgartner M., de Simone V., Obermaier B., Maché R., Mueller M.,
 RA Kreis M., Delzeny M., Puigdomenech P., Watson M., Schmidtheini T.,
 RA Reichert B., Portetelle D., Perez-Alonso M., Boutry M., Bancroft I.,
 RA Vos P., Hohnselt J., Zimmermann W., Wedler H., Ridley P.,
 RA Langham S.-A., McCullagh B., Bilham L., Robben J.,

RA Van der Schuren J., Grymonprez B., Chuang Y.-J., Vandenbussche F.,
 RA Braeken M., Wellens I., Voet M., Bastiaens I., Aert R., Defoor E.,
 RA Weltegenngger T., Bothe G., Ramsperger U., Hilbert H., Braun M.,
 RA Holzer E., Brandt A., Peters S., van Staveren M., Dirke P.,
 RA Moellman P., Klein lankhorst R., Rose M., Hauf J., Koetter P.,
 RA Berner S., Hempel S., Feldpausch M., Lamberth S., Van den Daele H.,
 RA De Keyser A., Buyschaert C., Gielen J., Villarroel R., De Clercq R.,
 RA Van Montagu M., Rogers J., Cronin A., Quail M., Bray-Allen S.,
 RA Clark L., Doggett J., Hall S., Kay M., Lennard N., McIay K., Mayes R.,
 RA Pettit A., Rajandream M.A., Lyne M., Benes V., Reckmann S.,
 RA Borikova D., Bloecher H., Scharfe M., Grimm M., Loehert T.-H.,
 RA Dose S., de Haan M., Maarse A.C., Schaefer M., Mueller-Auer S.,
 RA Gabel C., Fuchs M., Fartmann B., Granderath K., Dauner D., Herzl A.,
 RA Neumann S., Agirion A., Vitale D., Liguori R., Piravani E.,
 RA Massenot O., Quigley F., Clabaud G., Muendlin A., Felber R.,
 RA Schnabl S., Hiller R., Schmidt W., Lecharny A., Aubourg S.,
 RA Cheifor F., Cooke R., Berger C., Montfort A., Casacuberta E.,
 RA Gibbons T., Weber N., Vandenbol M., Barges M., Terol J., Torres A.,
 RA Perez-Perez A., Fumelle B., Bent E., Johnson S., Jacou D., Jesse T.,
 RA Heijnen L., Schwarz S., Scholler P., Heber S., Francis P., Bielek C.,
 RA Fishman D., Haase D., Lemcke K., Mewes H.-W., Stocker S.,
 RA Zaccaria P., Bevan M., Wilson R.K., de la Bastide M., Habermann K.,
 RA Parnell L., Dedhia N., Gnoj L., Schutz K., Huang E., Spiegel L.,
 RA Sekhon M., Murray J., Sheet P., Cordes M., Abu-Threideh J.,
 RA Stoneking T., Kallio J., Graves T., Harmon G., Edwards J.,
 RA Lattelle P., Courtney L., Cloud J., Abbott A., Scott K., Johnson D.,
 RA Minx P., Bentley D., Fulton B., Miller N., Greco T., Kemp K.,
 RA Kramer J., Fulton L., Mardis E., Dante M., Pepin K., Hillier L.,
 RA Nelson J., Speth J., Ryan E., Andrews S., Giesel C., Layman D.,
 RA Du H., Ali J., Berghoff A., Jones K., Dione K., Cotton M., Joshi C.,
 RA Antoniou B., Zidanic M., Strong C., Sun H., Lamar B., Jordan C.,
 RA Ma P., Zhong J., Preston R., Vill D., Shekher M., Matero A., Shah R.,
 RA Swaby I.K., O'Shaughnessy A., Rodriguez M., Hoffman J., Tili S.,
 RA Granat S., Shohdy N., Hasegawa A., Rodriguez M., Hoffman J., Tili S.,
 RA Chen E., Maira M., Martienssen R., McCombie W.R.;
 RT "Sequence and analysis of chromosome 4 of the plant Arabidopsis
 thaliana.";
 RL Nature 402:769-777(1999).
 CC -1- FUNCTION: MIGHT BE INVOLVED IN THE BIOSYNTHESIS OF STEROIDS
 CC IMPORTANT FOR THE POLAR ELONGATION OF CELLS DURING DEVELOPMENT.
 CC -1- SUBCELLULAR LOCATION: Membrane-bound. Endoplasmic reticulum
 CC (Potential).
 CC -1- TISSUE SPECIFICITY: UBIGINOUS.
 CC -1- SIMILARITY: BELONGS TO THE CYTOCHROME P450 FAMILY.
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC
 DR EMBL: AB008097; BA37167.1; -
 DR EMBL: AL161589; CAB80304.1; ALT_INIT.
 DR EMBL: Z99708; CAB16850.1; ALT_INIT.
 DR EMBL: AL022141; CA18139.1; ALT_SEQ.
 DR InterPro: IPR001128; Cytochrome_P450.
 DR Pfam: PF00067; P450; 1.
 DR PRINTS: PR00385; P450.
 DR PROSITE: PS00086; CYTOCHROME_P450; 1.
 KW Oxidoreductase; Monooxygenase; Transmembrane; Heme;
 KW Endoplasmic reticulum; Multigene family.
 FT TRANSMEM 4 24
 FT BINDING 463 463 HEME (BY SIMILARITY).
 FT CONFLICT 45 45 F -> L (IN REF. 2).
 SQ SEQUENCE 524 AA: 59389 MW: 550578908BDDP2722 CRC64;
 Query Match 32.2%; Score 864; Length 524;
 Best Local Similarity 35.4%; Pred. NO. 1.3e-44;
 Matches 186: Conservative 96: Mismatches 165: Indels 78: Gaps 10;

QY 24 FLIL-----LKRNRKRN-----LPPGSGMPFJGTYGKPYATTATLGDGFMDOH 56

```

Db 30 FLVLAAGILLRPWLFRNRKSKTKDGEDEENEEKKKMIPNGSLGMPVIGETLNFAC 89
OY 57 ---KPTATLTGDMQOHVSKYKIRSNLFGPTIVSADAGNRLTONEGLFECSYP 113
Db 90 YSRPPT-----FMKRSKSLGKVRKTNIGTPILISTDAENKVLQMHGNTFPAYP 143
OY 114 RSIGLILGKMSMLVLVGDHMRDRSISLNFSLHARLTILKDVREHRTFLDSMOONS 173
Db 144 KSITTELGNSLISLNGPHOKRHLTILGAFRPHLKDRTRDIEASVITLASMOLPL 203
OY 174 FSAODEAKKFTNLAHAKHMSMDPGEETDQKKEYVTMKGVASAPLPGTAHKAID 233
Db 204 VHVQDEIKMTFEILVKVLMSTSPG-EDNMILKLEFEFEIKGLICIPKPGTRLVSK 262
OY 234 SRATILKFERKMEKKLIDKEEDQEEVEYK-----EDEAMSKSDHVRKORTDOLL 287
Db 263 AERELIKMYKAYVEEQVAMTTSPANDVAVDLKDGSGSEKOPSDFVSGK----- 315
OY 288 GMYLKHSLNSTEOILDLILSLFAGHETSSVAIALAIFLOACPRAVEELREHLEIARA 347
Db 316 -----IYEMIPGEETMPTAMTLAVKFLSDMPVALAKLVEENMEKKR 358
OY 348 KEELGESELNMDYKMDTQCYINETLRLGNVRFLLHRAKALKDVRKYGYDIPSGKVL 407
Db 359 KLEIGE-EYKWDYMSLFTQVYNITELTMANINGVMRKALDVEIKGLIPKGCYLA 417
OY 408 VLSAVHLNRSYDOPNLFNPMRMOQONNGASSSGSFSFTWGNVNPFGGPRLCAGSEL 467
Db 418 STSYHMDDIYDNPQFDPMDRDRINGSANSSIC-----FPPGGGRLCGLEL 468
OY 468 ALEMAVFIHLVLKFNWELADDOFAPFPVDFPGLPIRVSRI 512
Db 469 SKLEISIFLHLVTRYSW-TAEDEIVSPFYKMKRRLPIRVATV 512

RESULT 3
CP85_LYCES STANDARD: PRT: 464 AA.
AC 043147;
DT 15-DEC-1998 (Rel. 37, Created)
DT 15-DEC-1998 (Rel. 37, Last sequence update)
DT 15-DEC-1998 (Rel. 37, Last annotation update)
DE Cytochrome P450 85 (EC 1.14.-.-) (Dwarf protein).
GN CYP85 OR D.
OS Lycopersicon esculentum (Tomato).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
OC Asteridae; euasterids I; Solanales; Solanaceae; Solanum.
OX NCBI_TaxID=4081;
RN [1]
RN SEQUENCE FROM N.A.
RC STRAIN=CV. GCR758;
RX MEDLINE=9626705; PubMed=8672892;
RA Bishop G.J., Harrison K., Jones J.J.G.D.:
RT "The tomato Dwarf gene isolated by heterologous transposon tagging
RT encodes the first member of a new cytochrome P450 family.";
RT Plant Cell 8:959-969(1996).
CC -i- SIMILARITY: BELONGS TO THE CYTOCHROME P450 FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation-
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL: U54770; AAB17070.1; -
CC InterPro: IPR001128; Cytochrome_P450.
CC DR Pfam: PF00067; P450; 1.
CC DR PRINTS: PR00385; P450.
CC DR PROSITE: PS00086; CYTOCHROME_P450; 1.

```

```

KW Oxidoreductase; Monooxygenase; Membrane; Heme.
FT BINDING 414 414 HEME (BY SIMILARITY).
SQ SEQUENCE 464 AA; 53706 MW; DZB21AAB7B14E94 CRC64;
Query Match 28.4%; Score 761.5; DB 1; Length 464;
Best Local Similarity 33.4%; Pred. No. 1,4e-38;
Matches 170; Conservative 90; Mismatches 188; Indels 61; Gaps 10;

OY 12 LLLPSELILLFLILILKRN--RKTFNLPFGKSGMPETIGLYKPTATTLG-DEM 68
Db 5 LILFSFFGLICIFLTALRWNOVKYNNKRNPPGTMGPLTGETTEFLK-----LGPSFM 58
OY 69 OOHVSKGYKTYRSNLFGEPTIVSADAGNRLTONEGLRECSYPSRISIGLILGKMSMLV 128
Db 59 KNQARIGSFFKSHIILGCPITVMSDELNRIILNENKGLVPGTPQSMIDLIGCNTAAV 118
OY 129 VGDHMRDRSISLNFSLHARLTILKDVREHRTFLDSMOONSIFSADAEAKFTENLM 188
Db 119 NSAHKYMKGALLSLISPTMIRQDLPKIDEPKMSHLTN-DNKVIDIQETKNMFLSS 177
OY 189 AKHIMSMDPEEET---EQLKKEYVTMKGVASAPLPGTAHKALOSRTATILKTERK 245
Db 178 LKQI---AGIESTSLAQEFMSSEFNVLGTLSPILPMTNHRGFQARKIYVNLRTL 233
OY 246 MEERKLDIKREDEQEEVEKTEDEAEMSKSDHVRKORTDDLLGMYLKH-----NLSTEOI 301
Db 234 IEERR-----ASKELQHDMLGTLNMEATREKLTDEM 266
OY 302 LDLLSLFAGHETSSVAIALAIFLOACPRAVEELREHLEIARAKKEGESELNMDY 361
Db 267 IDLITLIVSGYEFTVSTSMVAVKYLLHDHFKLELELKEHMAIREKKP--EDPIDYNDY 324
OY 362 KKMDTQCYINETLRLGNVRFLLHRAKALKDVRKYGYDIPSGKVLPIVSAVHLNRSYD 421
Db 325 RSMRFTRAVTLFSRLATVINGVLRKTQDMEINGVLIIPGWRYYVTRRLNDPRLYPD 384
OY 422 PNLFPNPMOQONNGASSSGSFSFTWGNVNPFGGPRLCAGSELAKLEMAVFIHLVL 481
Db 385 PYSFNPMRWMDKS-----LEHQNSELVFPFGGTRGCPKRELGVATISFTLHFTV 433
OY 482 KFNWELADDOFAPFPVDFPGLPIRVS 510
Db 434 KYRWEEIGDKLKMFPVEADNGLIRIVS 462

RESULT 4
C883_ARATH STANDARD: PRT: 490 AA.
AC 023051;
DT 15-DEC-1998 (Rel. 37, Created)
DT 15-DEC-1998 (Rel. 37, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Cytochrome P450 88A3 (EC 1.14.-.-).
GN CYP88A3 OR ATIG05160 OR YUP8H12.23.
OS Arabidopsis thaliana (Mouse-ear cress).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;
OC eurosids II; Brassicales; Brassicaceae; Arabidopsis.
OX NCBI_TaxID=5702;
RN [1]
RN SEQUENCE FROM N.A.
RC STRAIN=CV. Columbia;
RX MEDLINE=21016719; PubMed=11130712;
RA Theologis A., Ecker J.R., Palm C.J., Federspiel N.A., Kaul S.,
RA White O., Alonso J., Altati H., Arujo R., Bowman C.L., Brooks S.Y.,
RA Buehler E., Chan A., Chao Q., Chen H., Cheuk R.F., Chin C.W.,
RA Chung M.K., Conn L., Conway A.B., Conway A.R., Creasy T.H., Dewar K.,
RA Dunn P., Elgu P., Feldblum T.V., Feng J.-D., Fong B., Fujii C.Y.,
RA Gali J.E., Goldsmith A.D., Haas B., Hansen N.F., Hughes B., Hultzer L.,
RA Hunter J.L., Jenkins J., Johnson-Hopson C., Khan S., Khaykin E.,
RA Kim C.J., Koo H.L., Kremenská I., Kuritz D.B., Kwan A., Lam B.,
RA Langin-Hooper S., Lee A., Lee J.M., Lenz C.A., Li J.H., Li Y.-P.,
RA Lin X., Liu S.X., Liu Z.A., Luros J.S., Maiti R., Marziani A.,

```

RA Miltischer J., Miranda M., Nguyen M., Nierman W.C., Osborne B.I.,
 RA Pei G., Peterson J., Pham P.K., Rizzo M., Rooney T., Rowley D.,
 RA Sakano H., Salzberg S.L., Schwartz J.R., Shinn P., Southwick A.M.,
 RA Sun H., Tallon L.J., Tambunga G., Toriumi M.J., Town C.D.,
 RA Uterback T., Van Aken S., Vaysberg M., Vysotskaya V.S., Walker M.,
 RA Wu D., Yu G., Fraser C.M., Venter J.C., Davis R.W.,
 RT "Sequence and analysis of chromosome 1 of the plant Arabidopsis
 RT thaliana."
 RL Nature 408:816-820(2000).
 CC -1- SIMILARITY: BELONGS TO THE CYTOCHROME P450 FAMILY.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 CC EMBL: AC000098; AAB71462.1; -.
 DR HSSP: P14779; 1JPZ.
 DR InterPro: IPR001128; Cytochrome_P450.
 DR Pfam: PF00067; P450; 1.
 DR PRINTS: PR00385; P450.
 DR PROSITE: PS00086; CYTOCHROME_P450; 1.
 KM Oxidoreductase; Monooxygenase; Transmembrane; Heme.
 FT TRANSMEM 6
 FT BINDING 439 439 HEME (BY SIMILARITY).
 SQ SEQUENCE 490 AA; 56409 MW; 7ED8CD7A8864D288 CRC64;
 Query Match 21.6%; Score 579.5; DB 1; Length 490;
 Best Local Similarity 28.6%; Pred. No. 1.1e-27;
 Matches 133; Conservative 107; Mismatches 192; Indels 33; Gaps 9;
 QY 34 KTFENLPGRSGKPELGETIGYLYKPYATATLGDPMQOHVSKYK-1YRSNLPGEPTIVS 91
 DB 40 ENHYLPRLPGDLGMPFIGNMLSPFLAFKTSDDSTRLIRKYGKGIYKAMFENPSIIY 99
 QY 92 ADAGLNFLIIONEGRLFCESYPRNSIGILKWSMLVLVGDHMRISILNLSHARLRT 151
 DB 100 TTSDTGCRRLVLTDD-AFKPMPSTMLLGRKSEVGSISFEHKKRLRLTAPVNGHEALS 158
 QY 152 ILKDYVRHTLFLVDSQMSISFAODEAKKFTFNLMAKHMSMDPGEETEOAKKEVYT 211
 DB 159 TYIPYEENVITVLDKTKMGEFEFTLTKLTPRII-MYIFLSESENVADALEREYTA 217
 QY 212 FMGVVASAPLPLPGTAVYHAKLQSRATILKEFKMEERKIDKEEDOESEKTEDEAEM 271
 DB 218 LNVGVRAAMVNIQGFARHAKRLAKRTIYAFQSIYVTRR-----NORKNIILSNKKDML 271
 QY 272 SKSDHVRKORTDLDGLVNLKSHNSLSTEQIIDLILSLFAGHETSVAIALAIFFLQACP 331
 DB 272 DNLNLNVK---DEED-----GKTLWDEETIDVLYMLNAGHESSGHTIMWAVFLQENP 320
 QY 332 KAVEELREHLELAKAKKEGESELMNDVKKMDFTQCIVNETIRLGNVAFRLHKKALKD 391
 DB 321 EVLQRAKAEQEMILKSRPE-GQKGLSLKEFKRMEFISQVAVDETALRTVTFSLTAERAKTD 379
 QY 392 VRYKGDIPSGKRVLPVIAVHLDNSRYDQNLFNPMRMOQNNNGASSGSGSFTWGN 451
 DB 380 VEMNGVILPKGMVLTWFRVYHIDPEVFPDRKEDPARW---DNG-----FVPRAGA 428
 QY 452 YMPFGGPRICAGSELAKLEMAVFIHHLVLKFMWELADDDPPAF 496
 DB 429 FLPEFGAGSHLCPGNDLAKLEISIFLHFLKLYQVKRSNPPCPVAY 473

DT 15-DEC-1998 (Rel. 37, Last annotation update)
 DE Cytochrome P450 26 (EC 1.14.-.-) (Retinoic acid-metabolizing
 DE cytochrome) (P450RA1) (Retinoic acid 4-hydroxylase).
 GN CYP26.
 OS Brachydanio rerio (zebrafish) (Danio rerio).
 OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
 OC Actinopterygii; Neopterygii; Teleostei; Ostariophysi; Cypriniformes;
 OC Cyprinidae; Danio.
 OX NCBI_Taxid=7955;
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=97094702; PubMed=8939936;
 RA White J.A., Guo Y.-D., Baetz K., Beckett-Jones B., Bonasoro J.,
 RA Hsu K.E., Dilworth F.J., Jones G., Petkovich M.;
 RT "Identification of the retinoic acid-inducible all-trans-retinoic
 RT acid 4-hydroxylase".
 RL J. Biol. Chem. 271:29922-29927(1996).
 CC -1- FUNCTION: PLAYS A KEY ROLE IN RETINOIC ACID METABOLISM. ACTS ON
 CC RETINOIDS, INCLUDING ALL-TRANS-RETINOIC ACID (RA) AND ITS
 CC STEROISOMER 9-CIS-RA. CAPABLE OF 4-HYDROXYLATION; HYDROXYLATION.
 CC RESPONSIBLE FOR GENERATION OF SEVERAL HYDROXYLATED FORMS OF RA,
 CC INCLUDING 4-OH-RA AND 4-OXO-RA.
 CC -1- SUBCELLULAR LOCATION: Membrane-bound. Endoplasmic reticulum.
 CC -1- INDUCTION: BY RETINOIC ACID.
 CC -1- SIMILARITY: BELONGS TO THE CYTOCHROME P450 FAMILY.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 CC EMBL: U68234; AAC60045.1; -.
 DR ZFIN: ZDB-GENE-980415-44; CYP26.
 DR InterPro: IPR001128; Cytochrome_P450.
 DR Pfam: PF00067; P450; 1.
 DR PRINTS: PR00385; P450.
 DR PROSITE: PS00086; CYTOCHROME_P450; 1.
 KM Oxidoreductase; Monooxygenase; Membrane; Heme; Microsome;
 KW Endoplasmic reticulum.
 FT BINDING 438 438 HEME (BY SIMILARITY).
 SQ SEQUENCE 492 AA; 56281 MW; F0471435B2F30509 CRC64;
 Query Match 18.8%; Score 503; DB 1; Length 492;
 Best Local Similarity 27.6%; Pred. No. 3.8e-23;
 Matches 144; Conservative 95; Mismatches 216; Indels 66; Gaps 14;
 QY 7 HTLLPLLLPSLLSLFL-----ILKRRNRKTRPNLPGRSGKPELGETIGYLYK 58
 DB 4 YTLWTFELCTIVLPALLFLAALKWEMIMIRVDPNCRSPRLPGTMGLPFIETQLI-- 61
 QY 59 YTTATLDPMQOHVSKYKIRSNLFGEPTIVSADAGLNFLIIONEGRLFCESYPRNSIG 118
 DB 62 ---LQRRKFLPKRKRQICITKTHLFGNPTVYKMGADNVKQLDLDEHKLVSYQMPASVPT 118
 QY 119 ILGKWSMLVLVGDHMRD-----MRSISINFLSHARLRTILKDYVRHTLFLVDSW-QQNS 172
 DB 119 ILGSDTILSNVGVQHKNNKKAIMRAFSDALEH-----YIPVIOQVKSALQENLQKDS 172
 QY 173 IFSADDEAKKFTFNLMAKHMSMDGEET--EOLKKEVYTFMKGVASAPLPLPGTAVYK 230
 DB 173 CVLVPEPKKMLFRIAMRILGLFEEQIKTDEQELVEAFEEEMIKMLFSLPDPVPSGLYR 232
 QY 231 ALQSRATILKEFKERKIDKEEDOESEKTEDEAEMSKSHVAKQRTDLDGLV 290
 DB 233 GLRAR---NFIHSIEENIKKKIQQDDNENQKTKDLQL-----IENSRSD----- 279
 QY 291 LKHSNLSLSTEQIIDLILSLFAGHETSVAIALAIFFLQACPVAEELREHLEIARAKE 350
 DB 280 ----PFLQAMKKEATLFLPGCHETTAATATSIYVFLQNLNTEVQKVAEE-----VQEKVE 331


```

QY 351 LG-----ESLNDNDYKRMDFQCVINETLRLGNVFLHKKALDKVRYKGYDIPSGMKVL 406
DB 332 KMATYRPGKLSMELLDQLKYTCGVICKEITLRINPPVGGFRVAKTFELNGLYQPKGNVI 391
QY 407 PVIASVHLDNSRYDQNLNPNRMQOONNKGASSSGSSTGMNNYMPFGGRLCAGSE 466
DB 392 YSICDTHADVADVPNKEEFQPERFM-----SKGEDGSRF--NYIPIGGSGRMCGVKE 442
QY 467 LAKLEMAVFIHHLVLFKFNWLEDDQ-----PAPFVDPDP 502
DB 443 FAKVLKILFVLELTOHCNWLISNGPMTKGTPIITYVDNLP 483

RESULT 6
C881_MAIZE
ID C881_MAIZE STANDARD: PRT: 519 AA.
AC 043246:
DT 15-DEC-1998 (Rel. 37, Created)
DT 15-DEC-1998 (Rel. 37, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Cytochrome P450 88A1 (EC 1.14.-.-) (DWARF3 protein).
GN CYP88A1 OR D3.
OS Zea mays (Maize).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACC clade;
OC Panicoideae; Andropogoneae; Zea.
OC NCBI_TaxID=4577;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-CV. B73.
RA MEDLINE=96004534; PubMed=7549486;
RX Winkler R.G., Helentjaris T.;
RT "The maize Dwarf3 gene encodes a cytochrome P450-mediated early step
RT in gibberellin biosynthesis.";
RL Plant Cell 7:1307-1317(1995).
CC -1- PATHWAY: gibberellin biosynthesis: early step.
CC -1- TISSUE SPECIFICITY: EXPRESSED IN ROOTS, DEVELOPING LEAVES, THE
CC VEGETATIVE MERISTEM, AND SUSPENSION CULTURE CELLS.
CC -1- SIMILARITY: BELONGS TO THE CYTOCHROME P450 FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL: U32579; AAC49067.1; -
DR InterPro: IPR001128; Cytochrome_P450.
DR Pfam: PF00067; P450; 1.
DR PROSITE: PS00086; CYTOCHROME_P450; 1.
KW Oxidoreductase; Monooxygenase; Transmembrane; Heme.
FT TRANSMEM 1 21
FT BINDING 466 466 HEME (BY SIMILARITY).
SQ SEQUENCE 519 AA; 57906 MW; 0F8977A024316D95 CRC64;

Query Match 18.3%; Score 491.5; DB 1; Length 519;
Best Local Similarity 28.2%; Pred. No. 2e-22;
Matches 129; Conservative 91; Mismatches 189; Indels 49; Gaps 11;

QY 36 RFLNPPGKSGWFLGTYLKYPTATLTGDFMQOYKSKYK--ITRSNLFGEPTI-VSA 92
DB 69 RARLPREGMGWPLVGVGMALFAFKSKGKDAFLASVYRRGRKGVRSFESSPYLVTT 128
QY 93 DAGLNFILIONEGRLFECSYPSISGILKWSMLVLVGMHMRMSISLNLFLSHARLTI 152
DB 129 ASGCKKVLMDDDA--FVTGMPKATVALVPRSEFVAPYDEHRRIRKRLTAPINFALTG 186
QY 153 LAKDVERHLLFLVDSW-QQNSIFSADDEAKKTFENMAKHIMSDGCEETEDLQKEYVT 211
DB 187 YLPFDITVTVSSLRAMADHGSVEFTLELRMTFKTIIVQ-IFLGADQATTRALERSTYE 245

```

```

QY 212 FMKGVSAPIPLPCTAYAKALOSRATILKFIERRMEERKLDIKE-----EDOEEREYKT 265
DB 246 LNYGMRMAINLPFAIRGALRRARRRLVAVLQGVLDERRARARAGVSGGYDMDRLEA 305
QY 266 EDEAEKMSKSDHVRKQRTDDELGVNLKHSNLSTQIIDLILSLPAGHETSSVAIALIF 325
DB 306 QDE-----RGRHLDLDD-----EILIVLVYMINAGHSSGHTMMATV 342
QY 326 FLQACPKRAVELBREHLEINARAKKELGSELMNDYKMDTQCVINETLRLGNVRLH 385
DB 343 FLQENPMPFARAKQDEALMRSIPS-SQKGLTLDFKRMETLSOVIDETLRLVNTSPVSF 401
QY 386 RKALKDVRKGYDIPSGMKVLPVISAHLNDSRYDQNLNPNRMQOONNKGASSSGSGSE 445
DB 402 RQATRDVFNVGYLPLPKMKVQLWLRVSHMDQVYPPDPKFPDSRME-----GHSRACTF 456
QY 446 STWGNNTMPFGGPRCLCAGSELAFLMAVFIHHLVLFK 483
DB 457 -----LAFGLGARLCPGNDLAKLEISVFLHFLGLGY 487

RESULT 7
C120_STYNY3
ID C120_STYNY3 STANDARD: PRT: 444 AA.
AC 059990:
DT 16-OCT-2001 (Rel. 40, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Putative cytochrome P450 120 (EC 1.14.-.-).
GN CYP120 OR CYP OR SLR0574.
OS Synechocystis sp. (strain PCC 6803).
OC Bacteria; Cyanobacteria; Chroococcales; Synechocystis.
OC NCBI_TaxID=1148;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=96127529; PubMed=8590279;
RA Kaneko T., Tanaka A., Sato S., Kotani H., Sazuka T., Miyajima N.,
RA Sugitara M., Tabata S.;
RT "Sequence analysis of the genome of the unicellular cyanobacterium
RT Synechocystis sp. strain PCC6803. I. Sequence features in the 1 Mb
RT region from map positions 648 to 928 of the genome.";
RL DNA Res. 2:153-166(1995).
RN [2]
RP SEQUENCE FROM N.A.
RX MEDLINE=97061201; PubMed=8905231;
RA Kaneko T., Sato S., Kotani H., Tanaka A., Asamizu E., Nakamura Y.,
RA Miyajima N., Hikosawa M., Sugitara M., Sasamoto S., Kimura T.,
RA Hosouchi T., Matsuno A., Muraki A., Nakazaki N., Nanno K., Okumura S.,
RA Shimpou S., Takeuchi C., Wada T., Watanabe A., Yamada M.,
RA Tabata S.;
RT "Sequence analysis of the genome of the unicellular cyanobacterium
RT Synechocystis sp. strain PCC6803. II. Sequence determination of the
RT entire genome and assignment of potential protein-coding regions.";
RL DNA Res. 3:109-136(1996).
CC -1- SIMILARITY: BELONGS TO THE CYTOCHROME P450 FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL: D64003; BAA10496.1; -
DR InterPro: IPR001128; Cytochrome_P450.
DR Pfam: PF00067; P450; 1.
DR PROSITE: PS00086; CYTOCHROME_P450; 1.
KW Hypothetical protein; Oxidoreductase; Monooxygenase; Heme;
FT BINDING 391 391 HEME (BY SIMILARITY).
SQ SEQUENCE 444 AA; 50578 MW; 8F62A9EED3B54BDC CRC64;

```

Query Match	17.9%	Score 479.5	DB 1	Length 444
Best Local Similarity	29.7%	Pred 20.85e-22		
Matches 137	Conservative 88	Mismatches 167	Indels 69	Gaps 13
QY	39	LPFGKSGMPFGETIGYLRKPYATVTLGFGMOOHASKYKIRYSNLFGEPTIVSADAGNR	98	
DB	13	IPGDFGLPWVGETELINFLND-----GDFGKKRQOQFPIKTRLFQKNVYFISALANR	66	
QY	99	FILNKGFLFECSIPRSITGGILGKWSMLVYGDHMRDRSISLNFSLHRL-RTI-LLK	155	
DB	67	FLFKEDETFOATWPLSTRILLGPNALATOMGEIHSRRKI----LYQAFLPRTIDSYLP	122	
QY	156	DVERHTLEFLWSMOONSIFSNQDEAKKRTENLMAKHIMSMPGEF--ETDQKKEYVTM	213	
DB	123	KMDGIQVQYTLQMGANENVITWYPOLRRTFVVAATLEW----GEKVSQNPQLFPWFETYL	178	
QY	214	KGVSAPLNLPGTAYHKAQSRATILKFIKKMEERKIDKEEDQEEVEVTEDEAEKSK	273	
DB	179	QGLSPFLPRLNTLEFGKSQRRALL-----AELK	209	
QY	274	SDHVRKQF--TDDDLGMVL-----KHSNLTQDILDLILSLFAGHETSSVAIALATF	326	
DB	210	IKKARQOQPSSEEDALGILLIARDDNNQPLSLPELKDQILLFLFAGHETTLTALSFCLL	269	
QY	327	LQACPKNAEELREHLEIARAKKELGESELMWDDYKKKDFQCYINETLRKGNVYRFLHR	386	
DB	270	LQSHSDIRERKQF-----QNKQLDSQ-ELTAETLTKKMYLDQVYLGVELRILPPVGGGR	323	
QY	387	KALDKVARKYGDIQSGMKVLEPVIASVHLDNSRYOQPLNFPMRMOQONNGASSSGSFS	446	
DB	324	ELIDQCGQGHFKGMVLSQISQIOTHADPDLTYDPEKFDERTPD---GSATINPFA	380	
QY	447	TWGNMYDFGGGPRICAGSELAKEMAVFIHHLVLEKNMEL	487	
DB	381	-----HVPFGGGLRECLGKEFAFLKMLFATRLIQGFDTWL	416	

RESULT 8

CP26_MOUSE STANDARD; PRT; 497 AA.

AC 055127; Q9RI1F4;

DT 15-DEC-1998 (Rel. 37, Created)

DT 15-DEC-1998 (Rel. 37, Last sequence update)

DT 15-JUN-2002 (Rel. 41, Last annotation update)

DE Cytochrome P450 26 (EC 1.14.-.-) (Retinoic acid-metabolizing

DE cytochrome) (P450RA1) (Retinoic acid 4-hydroxylase).

GN CYP26A1 OR CYP26 OR P450RA.

OS Mus musculus (Mouse).

OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.

OX NCBI_TaxID=10090;

RN [1]

RN SEQUENCE FROM N.A.

RP STRAIN=CH3;

RC MEDLINE=97392446; PubMed=9250660;

RA Fujii H., Sato T., Kaneko S., Gotoh O., Fujii-Kuriyama Y., Osawa K.,

RA Kato S., Hamada H.;

RT "Metabolic inactivation of retinoic acid by a novel P450

RT differentially expressed in developing mouse embryos.";

RL EMBO J. 16:4163-4173(1997).

RN [2]

RN SEQUENCE FROM N.A.

RP MEDLINE=9811312; PubMed=9442090;

RA Abu-Abed S.S., Beckett B.R., Chiba H., Chithalen J.V., Jones G.,

RA Metzger D., Channon P., Petkovich M.;

RT "Mouse P450RA1 (CYP26) expression and retinoic acid-inducible retinoic

RT acid metabolism in F9 cells are regulated by retinoic acid receptor

RT gamma and retinoid X receptor alpha.";

RL J. Biol. Chem. 273:2409-2415(1998).

RN [3]

RN SEQUENCE FROM N.A.

RP TISSUE=Tooth;

RC MEDLINE=20515038; PubMed=11063033;

```

RA Patne C.T., Patne M.L., Snead M.L.;
RT "Identification of tufellin- and amelogenin-interacting proteins using
the yeast two-hybrid system."
RL Connect. Tissue Res. 38:257-267(1998).
RN [4]
RP SEQUENCE FROM N.A.
RC RA
RC TISSUE=Liver;
RA Strausberg R.;
RL Submitted (Aug-2001) to the EMBL/GenBank/DDBJ databases.
CC -I- FUNCTION: Plays a key role in retinoic acid metabolism. Acts on
retinoids, including all-trans-retinoic acid (RA) and its
stereoisomer 9-cis-RA. Capable of both 4-hydroxylation and 18-
hydroxylation. Responsible for generation of several hydroxylated
forms of RA, including 4-OH-RA, 4-oxo-RA, AND 18-OH-RA.
CC -I- SUBCELLULAR LOCATION: Membrane-bound. Endoplasmic reticulum.
CC -I- INDUCTION: BY RETINOIC ACID.
CC -I- SIMILARITY: BELONGS TO THE CYTOCHROME P450 FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation at
the European Bioinformatics Institute. There are no restrictions on
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (see http://www.isb-sib.ch/announce/
or send an email to license@isb.sib.ch).
CC -----
DR EMBL; Y12657; CAAT33206.1; -.
DR EMBL; AF115769; AAD17217.1; -.
DR EMBL; BC012673; AAH12673.1; -.
DR HSSP; P14779; 1JPZ.
DR MD; MG1:1096359; Cyp26a1.
DR InterPro; IPR001128; Cytochrome_P450.
DR Pfam; PF00067; P450; 1.
DR PRINTS; PR00385; P450.
DR PROSITE; PS00086; CYTOCHROME_P450; 1.
KW Oxidoreductase; Monooxygenase; Membrane; Heme; Microsome;
KM Endoplasmic reticulum.
FT BINDING 442 442 HEME (POTENTIAL).
FT CONFLICT 9 9 S-> T (IN REF. 3).
FT CONFLICT 154 154 L-> P (IN REF. 4).
FT CONFLICT 356 356 I-> T (IN REF. 4).
FT CONFLICT 492 492 Y-> H (IN REF. 4).
SQ SEQUENCE 497 AA; 56177 MW; 33B07DD7C9134471 CNC64;

Query Match 17.2% Score 461; DB 1; Length 497;
Best Local Similarity 28.4% Pred. No. 1.2e-20;
Matches 148; Conservative 78; Mismatches 231; Indels 64; Gaps 14;

QY 10 LPILLPSLSL-----SLFLFILK-----RRNRTRFNLPGKSGMPLEGTIGLK 57
Db ||||| :||| | | | | | | | | | | | | | | | | | | | | | | | | | |
3 LPALLASALCTFYVLPILFLAALKLMDLYCYSSSDRSCALPLPFGTMGFFPGETL---- 58

QY 58 PYATTTLGDWMQOHVSRYGKIYSNLSNGEPTIVSADAGLNLFITLNDGRRLFECYSPRSIG 117
Db ::::: :||| ||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
QY 59 -QWLVQRKFLOMKRRKYGFYKTHLPGRPIRVGMADVNRILLGEHRIVSYHMPPASVR 117
Db ::::: :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||

QY 118 GLIGKSMVLVCGMHDMSISLNFSHARKLTLLKDVERHTLFLVDMSQQ--NSTFS 175
Db ||||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
QY 118 TLTGAGCLSNLDHSSHQRKKVLMQAESREALQCVLV-TAEVYSCLCLEMLSCGERGL 176
Db ||||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||

QY 176 AODEAKRFETNLMAKHIMSDPG-----EEEOQLKEEVFMKGVNASAPNLPTAYHKA 231
Db ||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
Db 177 VYPEVKRLMRIRAKRIILGCEPERGAGEDEQQLVEAFEEETRNLFLSLPIDVPSGLRG 236

QY 232 LOSFAATLKFTERRME--RKLDIKEED---QEEEVKTDEADMKS SDHYRKORTDDL 286
Db ::::: :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
Db 237 VKARNLIHARTIEENIRAKIRRLQATEPDGCCKDALOLIEHSWE-----RGRLDMQ- 288

QY 287 LGHWLKRSNSTEDIIDLLISLFPAGHEITSVALALAIFPLQCPRAVEELREHLER 346
Db ||||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
Db 289 ---ALKGS-STE-----LLFGHETTASAATSILTYGLPYHVLQKRREEIKSKGL 335

XZ AKKELGSESLMWDDYKRMDFQCINTEFLRGVNVFRLLHKRALDVMYKKGDIIPSGKV 406

```

```

Db      336 LCKSNODNKLDMETLEOLKXIGYIKETLRLNPNPVGFRVAKTFELNGYQIPKGNVI 395
QY      407 PVISAVHLDSRYDQPLFNPMWRMQOONNASSGSSFSFTWGNMYNPFEGGRLCAGSE 466
Db      396 YSICDTHDVADIFTNKEEFNDRIVPHEDASRFS-----FIPFGGLRSCVCKE 446
QY      467 LAKLEMAVFTLHLVLFKNWELAE-----DDQPPAFPEVDP 502
Db      447 FAKILKIFVELARHCDQMLNGPPTMKTSPYIVPDNLP 487

RESULT 9
C26A_HUMAN
ID C26A_HUMAN STANDARD; PRT; 512 AA.
AC Q9NR63; Q9NP41;
DT 15-JUN-2002 (Rel. 41, Created)
DT 15-JUN-2002 (Rel. 41, Last sequence update)
DE Cytochrome P450 26A2 (EC 1.14.--) (P450RA1-2) (Retinoic-acid
  metabolizing cytochrome).
GN CYP26A2 OR P450RA12.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Retina;
RX MEDLINE=20300913; PubMed=10823918;
RA White J.A., Ramshaw H., Taimi M., Stangle W., Zhang A., Everingham S.,
RT Creighton S., Tam S.-P., Jones G., Petkovich M.;
RT "Identification of the human cytochrome P450, P450RA1-2, which is
  predominantly expressed in the adult cerebellum and is responsible for
  all trans-retinoic acid metabolism."
RT Proc. Natl. Acad. Sci. U.S.A. 97:6403-6408 (2000).
RL [2]
RP SEQUENCE FROM N.A.
RA Bourne S., Bauer C., Pape K., Jones T.;
RL Submitted (MAR-1999) to the EMBL/GenBank/DBJ databases.
CC -1- FUNCTION: Plays a key role in retinoic acid metabolism. Involved
  in the specific inactivation of all-trans-RA. Responsible for
  generation of several hydroxylated forms of RA, including 4-OH-RA,
  4-oxo-RA, and 18-OH-RA.
CC -1- ENZYME REGULATION: Has a preferred activity towards the following
  substrates: all-trans-RA > 9-cis-RA > 13-cis-RA.
CC -1- SUBCELLULAR LOCATION: Membrane-bound. Endoplasmic reticulum (By
  similarity).
CC -1- TISSUE SPECIFICITY: Highly expressed in brain, particularly in the
  cerebellum and pons.
CC -1- SIMILARITY: BELONGS TO THE CYTOCHROME P450 FAMILY.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
  between the Swiss Institute of Bioinformatics and the EMBL outstation-
  use by non-profit institutions as long as its content is in no way
  modified and this statement is not removed. Usage by and for commercial
  entities requires a license agreement (See http://www.isb-sib.ch/announce/
  or send an email to license@isb-sib.ch).
CC -----
DR EMBL: AF252297; AAF76003.1; -
DR EMBL: AC007002; AAF55576.1; -
DR MIM: 605207; -
DR InterPro: IPR001128; Cytochrome_P450.
DR Pfam: PF00067; P450_2.
DR PRINTS: PR00385; P450.
DR PROSITE: PS00086; CYTOCHROME_P450_1.
KW Oxidoreductase; Monooxygenase; Membrane; Heme; Microsome;
KW Endoplasmic reticulum
FT BINDING 441 441 HEME (POTENTIAL).
FT CONFLICT 68 68 Q -> QKCTLRTRVWMLQ (IN REF. 2).
SQ SEQUENCE 512 AA; 57512 MW; A06DID944E6726F CRC64;

```

```

Query Match 16.68; Score 446; DB 1; Length 512;
Best Local Similarity 26.98; Pred. No. 9.9e-20;
Matches 132; Conservative 86; Mismatches 219; Indels 54; Gaps 10;

QY 14 LPSLSLLFLILKLR-----NKRTRFLPPGKSGWPLGFTIGLYKRYATTL 64
Db 15 LAACLSVTLILLAVSOOLRMATYDKSCRLPIPGSMGFPLIGETGMLLGSG-- 71
QY 65 GDEMOQHVSKYKGYRNSLGEPTIVSADAGLNRFILQNGRLEFCSPRSIGILGKWS 124
Db 72 --FQSSRREKYGVNFKHLLGRPLIRYTGAEENRKILMGHHLYSTWEPSTRLLGPT 129
QY 125 MLYLVGDMDHMDMSISNPLSHARLTILIKDYERHTLFVLDWMQON-SIFSADQAKKF 183
Db 130 VSNISGDIHNRKRVKSKIPSHALESYLPK-1QLVLDPLRAMSSHPEAINVYQEOKL 188
QY 184 TFMIMAKHIMSMDPGEETEQLKREYVTEFKGVASAPLNPGRYHKAOSRATILKFE 243
Db 189 TFMARIVLIGFSIPEDLGLHEVYQGFVDFVSLFVDFPSGRRGIQARQILOGLE 248
QY 244 RKMERRLDIKEDDEEEVKTDEAEMSKSDHVRRKQRTDDLLGWILKHSNSTEQIID 303
Db 249 KAIREK----LQCTGQDYLADLLIESKEHKE-----WFMQELKD 288
QY 304 LILSLFAGHETSVAIALAIFLOACPKAVEELREBHETIARAKKELG-----ESELN 357
Db 289 GTLELLEFAVYATTAASASTSLIMQLKHPVLEKLRDE---LRAGHILHSGGCPGRTLR 344
QY 358 WDDYKRMDFLOCYINFTLRIGNVREFLRKALKDVAHYKGYDPSGKVLPIYSAVHLDS 417
Db 345 LDTLSLRILDCYIKRYMRLFTYISGKYRVLDTFELDGFQIRPKGSVMKSTINDDTAP 404
QY 418 RYDQPLFNPMWRMQOONNASSGSSFSFTWGNMYNPFEGGRLCAGSELAKEMAVFTI 477
Db 405 VFEDVAVFDPDRFSQ--ARSEDKDGRF-----HYLPFGGVYTCGLKHLAKFLVAV 456
QY 478 HLVLKFNWELA 488
Db 457 ELASTSRFEELA 467

RESULT 10
CP26_XENLA
ID CP26_XENLA STANDARD; PRT; 492 AA.
AC Q93323;
DT 15-JUN-2002 (Rel. 41, Created)
DT 15-JUN-2002 (Rel. 41, Last sequence update)
DE Cytochrome P450 26 (EC 1.14.--) (Retinoic acid degrading enzyme
  CYP26) (XCPY26) (Retinoic acid converting enzyme) (RACE).
GN CYP26A1 OR CYP26.
OS Xenopus laevis (African clawed frog).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Amphibia; Batrachia; Anura; Mesobatrachia; Pipidae;
OC Xenopodinae; Xenopus.
OX NCBI_TaxID=8355;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=99077803; PubMed=9857192;
RA Hollemann T., Chen Y., Grunz H., Piejer T.;
RT "Regionalized metabolic activity establishes boundaries of retinoic
  acid signaling."
RL EMO J. 17:7361-7372(1998).
CC -1- FUNCTION: Plays a key role in retinoic acid metabolism (By
  similarity).
CC -1- SUBCELLULAR LOCATION: Membrane-bound. Endoplasmic reticulum (By
  similarity).
CC -1- INDUCTION: By retinoic acid.
CC -1- SIMILARITY: BELONGS TO THE CYTOCHROME P450 FAMILY.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
  between the Swiss Institute of Bioinformatics and the EMBL outstation -

```


QY 287 LGWLKHS-----NISTEOLIDLILSLFAGHETSVAIALAIFFLQACRAVEELREH 341
 DB 266 LQLMHEHTOGNGBDLNNQELKESATLILFGSHETASANSILAFILG-----HH 315
 QY 342 LEIARAKEL-----GESELMNDVYKMDFTOCVINEIRLGNVAFELRKALKD 391
 DB 316 DVLOKVRKELOLKGLSLGPNQEROLNMEFLEQLKTYGCVIKETIRLSPVPGGRIRALKT 375
 QY 392 VRYKGYDIPSGWKLPYISAVHLNDSKRDQPNLFNPMWOOQNNGASSGSGSPSTWGN 451
 DB 376 LEIINGYDIPKGMWNYISICDTHDVADELFTDKDEFNPDPRFM-----SPSPEDSSRPS-----426
 QY 452 YMPGGPRLCAGSELAKLEMAVFIHHLVLKFNELAEADQ-----PFAPPEVDFP 502
 DB 427 FIPGGGLRSCGCVGEFAKVLKIFTVELARSCMQLNGPPTMKTGTVYVVDNLP 482

RESULT 12
 CP26_HUMAN STANDARD; PRT; 497 AA.
 AC 043174;
 DT 15-DEC-1998 (Rel. 37, Created)
 DT 15-DEC-1998 (Rel. 40, Last sequence update)
 DE 16-OCT-2001 (Rel. 40, Last annotation update)
 DE Cytochrome P450 26 (EC 1.14.-.-) (Retinoic acid-metabolizing
 CYP26A1 OR CYP26).
 GN CYP26A1 OR CYP26.
 OS Homo sapiens (Human).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
 OX NCBI_TaxID=9606;
 RN [1]
 RP SEQUENCE FROM N.A.
 RA MEDLINE-97373542; PubMed-9228017;
 RA White J.A., Beckelt-Jones B., Guo Y.-D., Dilworth F.J., Bonasoro J.,
 RA Jones G., Petkovich M., van der Burg B., van der Saag P.T.;
 RT "cDNA cloning of human retinoic acid-metabolizing enzyme (hP450RA1)
 RT identifies a novel family of cytochromes P450.";
 RL J. Biol. Chem. 272:18538-18541(1997).
 RN [2]
 RP SEQUENCE FROM N.A.
 RA MEDLINE-98380037; PubMed-9716180;
 RA Sonneveld E., van den Brink C.E., van der Leede B.M., Schalkes R.K.,
 RA Petkovich M., van der Burg B., van der Saag P.T.;
 RT "Human retinoic acid (RA) 4-hydroxylase (CYP26) is highly specific for
 RT all-trans-RA and can be induced through RA receptors in human breast
 RT and colon carcinoma cells.";
 RL Cell Growth Differ. 9:629-637(1998).
 RN [3]
 RP TISSUE SPECIFICITY.
 RX MEDLINE-99045433; PubMed-9826557;
 RA Trofimova-Griffin M.E., Juchau M.R.;
 RT "Expression of cytochrome P450RA1 (CYP26) in human fetal hepatic and
 RT cephalic tissues.";
 RL Biochem. Biophys. Res. Commun. 252:487-491(1998).
 CC -1- FUNCTION: Plays a key role in retinoic acid metabolism. Acts on
 CC retinoids, including all-trans-retinoic acid (RA) and its
 CC stereoisomer 9-cis-RA. Capable of both 4-hydroxylation and 18-
 CC hydroxylation. Responsible for generation of several hydroxylated
 CC forms of RA, including 4-OH-RA, 4-oxo-RA, and 18-OH-RA.
 CC -1- SUBCELLULAR LOCATION: Membrane-bound. Endoplasmic reticulum.
 CC -1- TISSUE SPECIFICITY: Highest levels in adult liver, heart,
 CC pituitary gland, adrenal gland, placenta and regions of the brain.
 CC -1- INDUCTION: By retinoic acid.
 CC -1- SIMILARITY: BELONGS TO THE CYTOCHROME P450 FAMILY.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>)

CC or send an email to license@isb-sib.ch.
 CC -----
 CC EMBL; AF005418; AAB88881.1; .
 DR HSSP; P14779; 1JPZ.
 DR Genew; HGNC:2603; CYP26A1.
 DR MIM; 602239; .
 DR InterPro; IPR001128; Cytochrome_P450.
 DR Pfam; PF00067; P450.1.
 DR PRINTS; PR00385; P450.
 DR PROSITE; PS00086; CYTOCHROME_P450; 1.
 DR Oxidoreductase; Monooxygenase; Membrane; Heme; Mitochondrion;
 KW Endoplasmic reticulum.
 FT BINDING 442 442 HEME (POTENTIAL).
 FT SEQUENCE 497 AA; 56162 MW; EAB684B24B2EAB3 CRC64;
 SO

Query Match 16.3%; Score 438; DB 1; Length 497;
 Best Local Similarity 27.08; Pred. No. 2.9e-19;
 Matches 142; Conservative 85; Mismatches 226; Indels 72; Gaps 15;

QY 10 LPLLLPSL-----LSLLFLTLK-----RRNRKTRNLPDPSGSMPLFGTIGYLK 57
 DB 3 LPALLASALCTFVLPPLFLFAIKIMDLVCVSGDRSCALPLDPGTGPPFFGFTL-----58
 QY 58 PYATTLGDPMQOHVSXGKIYSNLEGEPTIYSADGLNRFILQNGRLFECSYPSISIG 117
 DB 59 -QWVLOKRRKFLQKRRRYGFTYTHLEGRPTVRMGADVNRILLGDRLVSVMHPASVR 117
 QY 118 GILGKSMVLVGDMDRMSISLNEFSHARLR---TILKDYERHTLFLVDSMQQ--NS 172
 DB 118 TILGSGCLSNLHDSHQRKVKVIMRAFSRALCYVYITIEVGSS---LEOMLSGGER 173
 QY 173 IFSAODEAKKFTTNLAKKIHMSDP-----GEEETDLKKEYVTFMKGVASADLNLPGTA 227
 DB 174 GLIYDPEVKRLMRIMARILLGCEPLAGDSE--QOLVAFEMTRNLFSLIDVDFSG 232
 QY 228 YKRALSRATILFERKMEERLTDIKEED-----QEEETDLKKEYVTFMKGVASADLNLPGTA 282
 DB 233 LYRGKMARNLIHRIQONIRAKICGLRASAGGCCDAQDLLEHSE-----GGERL 285
 QY 283 DDDLLGWLKSLSSTEQILDLILSLFAGHETSVAIALAIFFLQACRAVEELREH 342
 DB 286 DMQ---ALKOS---STE-----LLFGSHETASANSILTYLGLYVHLQKVRDELK 331
 QY 343 ELIARAKELGESELMNDVYKMDFTOCVINEIRLGNVAFELRKALKDVRKYGYDIPSG 402
 DB 332 SKGLCKSMQNDKLDMEILLQKTYGCVIKETIRLNPVPGGRVAKLTFELNGYQIPKG 391
 QY 403 WKVLPYISAVHLNDSKRDQPNLFNPMWOOQNNGASSGSGSPSTWGNVMPGGPRLC 462
 DB 392 WNYIYSICDTHDVAELFTKKEENPDPRFMLPHEDSSRPS-----FIPGGGLRSC 442
 QY 463 AGSELAKLEMAVFIHHLVLKFNELAE-----DDPFAPPEVDFP 502
 DB 443 VKEPEAKLILKIFTVELARSCMQLNGPPTMKTGTVYVVDNLP 487

RESULT 13
 YRV2_CAEEL
 ID YRV2_CAEEL STANDARD; PRT; 520 AA.
 AC 027514;
 DT 01-NOV-1997 (Rel. 35, Created)
 DT 01-NOV-1997 (Rel. 35, Last sequence update)
 DE 30-MAY-2000 (Rel. 39, Last annotation update)
 DE Putative cytochrome P450 CYP13A5 (EC 1.14.-.-).
 GN CYP13A5 OR TIOB9.2.
 OS Caenorhabditis elegans.
 OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditioidea;
 OC Rhabditidae; Pelodierinae; Caenorhabditis.
 OX NCBI_TaxID=6239;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN-Bristol N2;
 RA Gardner A.;

RL Submitted (MAR-1995) to the EMBL/GenBank/DBJ databases.
 CC -1- FUNCTION: CYTOCHROMES P450 ARE A GROUP OF HEME-THIOLATE
 CC MONOOXYGENASES. THEY OXIDIZE A VARIETY OF STRUCTURALLY UNRELATED
 CC COMPOUNDS, INCLUDING STEROIDS, FATTY ACIDS, AND XENOBIOTICS.
 CC -1- SIMILARITY: BELONGS TO THE CYTOCHROME P450 FAMILY.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 DR EMBL; Z48717; CAAB8604.1; -;
 DR HSSP; P14779; IJPZ.
 DR WormPep; T10B9.2; CE01655.
 DR InterPro: IPR001128; Cytochrome_P450.
 DR Pfam: PF00067; P450; 1.
 DR PRINTS; PR00385; P450.
 DR PROSITE; PS00086; CYTOCHROME_P450; 1.
 DR Hypothetical protein; Oxidoreductase; Monooxygenase; Heme.
 FT BINDING 464 464 HEME (BY SIMILARITY).
 SQ SEQUENCE 520 AA; 59524 MW; 0B7B19E25B7ADFB3 CRC64;

 Query Match 12.18; Score 324.5; DB 1; Length 520;
 Best Local Similarity 23.68; Pred. No. 1,7e-12;
 Matches 120; Conservative 94; Mismatches 178; Indels 117; Gaps 20;

 QY 42 GKSGMPFLG-----ETIGLKPYATATLGDPMQOHVSKYGIYSNLFGEPTIYSAD 93
 DB 33 GPRGPFPGVYIHEDYENPGLK-----LGWTKE---YGIYGTGCEVETLIVSN 82
 QY 94 AGLNFTLQNGRLFECSYPSISGILG-----KMSMLVLY-GMHMDRISINFLSHA 147
 DB 83 P---EFVHEVFKQDFNFGKRTNPLOGDPNKNRAHLYSQGHMKRLRLTSPFSNK 139
 QY 148 RLRTIL-----LKQVRRHTLVLDSSQONSISFADDAKFTFNMAKHISMDEE 201
 DB 140 NLKRMSTVEETVELMRH---LDDASAKKAVLDLYOEFITDIIGRIAM---GOTE 192
 QY 202 TEOLKREYVFMKG-----VSAPLNPCTAVHK-----ALQSRATILKF 241
 DB 193 SLMFRRNMLPKVKGIFKDGKRLPPLVSGIFPIAGTMRREFPMRPSIQAPVDINSTEKA 252
 QY 242 IERKMERKLDIKE-----EDQ-----EVEEYKTEDEAKMSKSDHYRK 279
 DB 253 LNKRLRGRADEKAGIEPSEPODFIDFLDARANVDFEESALGKAKTEIAVD---K 309
 QY 280 QRTDDLLGWLKHSNLSTEQIIDLILSLFAGHETSSVAIALAIFLQACPAVEELRE 339
 DB 310 QLFDELDIG-----QLFVFLAGYDTALSLSSSYLLARHPEIQKLOE 354
 QY 340 EHLERARAKKEGSELNMDYKMDFTOCVINEETLGNVYFLH-RKALKDVRKGYD 398
 DB 355 E-----VRECNPEPTTFQISKLYMECVKEALMTPLASTVHNKKCKEETNVLGQ 408
 QY 399 IPGKMWLVPIYSAVHLDSRY-DQPLNFPWRMOQONNGASSGSGSFTGNNYPMFGG 457
 DB 409 IEKGTNVQVDTWTLTHDPKVMGEDANFERPERME-----SGDELFPYAKG-GYLPFGM 459
 QY 458 GPRLCAGSELAKLEMAVFIHHLVLFKNME 486
 DB 460 GPRICIGMLAMMEKKMLTHILKKYTFE 488

 RESULT 14
 ID YRYA_CABEL STANDARD; PRT; 518 AA.
 AC Q27519;
 DT 01-NOV-1997 (Rel. 35, Created)
 DT 01-NOV-1997 (Rel. 35, Last sequence update)
 DT 30-MAY-2000 (Rel. 39, Last annotation update)

DE Putative cytochrome P450 CYP13A7 (EC 1.14.-.-).
 GN CYP13A7 OR T10B9.10.
 OS Caenorhabditis elegans.
 OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditioidea;
 OC Rhabditidae; Peloderinae; Caenorhabditis.
 OX NCBI_TaxID=6239;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN-Bristol N2;
 RA Garner A.;
 RL Submitted (MAR-1995) to the EMBL/GenBank/DBJ databases.
 CC -1- FUNCTION: CYTOCHROMES P450 ARE A GROUP OF HEME-THIOLATE
 CC MONOOXYGENASES. THEY OXIDIZE A VARIETY OF STRUCTURALLY UNRELATED
 CC COMPOUNDS, INCLUDING STEROIDS, FATTY ACIDS, AND XENOBIOTICS.
 CC -1- SIMILARITY: BELONGS TO THE CYTOCHROME P450 FAMILY.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 DR EMBL; Z48717; CAAB8609.1; -;
 DR HSSP; P14779; IJPZ.
 DR WormPep; T10B9.10; CE01655.
 DR InterPro: IPR001128; Cytochrome_P450.
 DR Pfam: PF00067; P450; 1.
 DR PRINTS; PR00385; P450.
 DR PROSITE; PS00086; CYTOCHROME_P450; 1.
 DR Hypothetical protein; Oxidoreductase; Monooxygenase; Heme.
 FT BINDING 464 464 HEME (BY SIMILARITY).
 SQ SEQUENCE 518 AA; 58999 MW; CC04283EF87B9EAT7 CRC64;

 Query Match 12.08; Score 321.5; DB 1; Length 518;
 Best Local Similarity 24.08; Pred. No. 2.6e-12;
 Matches 120; Conservative 96; Mismatches 184; Indels 101; Gaps 20;

 QY 42 GKSGMPFLG-----ETIGLKPYATATLGDPMQOHVSKYGIYSNLFGEPTIYSAD 93
 DB 33 GPRGPFPGVYIHEDYENPGLK-----LGWTKE---YGIYGTGCEVETLIVSN 82
 QY 94 AGLNFTLQNGRLFECSYPSISGILG-----KMSMLVLY-GMHMDRISINFLSHA 147
 DB 83 P---EFVHEVFKQDFNFGKRTNPLOGDPNKNRAHLYSQGHMKRLRLTSPFSNK 139
 QY 148 RLRTI-----LKQVRRHTLVLDSSQONSISFADDAKFTFNMAKHISMDEE 201
 DB 140 NLKRMSTVEETVELMRH---LDDASAKKAVLDLYOEFITDIIGRIAM---GOTE 192
 QY 202 TEOLKREYVFMKG-----VSAPLNPCTAVHK-----ALQSRATILKF 241
 DB 193 SLMFRRNMLPKVKGIFKDGKRLPPLVSGIFPIAGTMRREFPMRPSIQAPVDINSTEKA 252
 QY 242 IERKMERKLDIKE-----EDQ-----EVEEYKTEDEAKMSKSDHYRK 279
 DB 253 LNKRLRGRADEKAGIEPSEPODFIDFLDARANVDFEESALGKAKTEIAVD---K 309
 QY 280 QRTDDLLGWLKHSNLSTEQIIDLILSLFAGHETSSVAIALAIFLQACPAVEELRE 339
 DB 310 QLFDELDIG-----QLFVFLAGYDTALSLSSSYLLARHPEIQKLOE 354
 QY 340 EHLERARAKKEGSELNMDYKMDFTOCVINEETLGNVYFLH-RKALKDVRKGYD 398
 DB 355 E-----VRECNPEPTTFQISKLYMECVKEALMTPLASTVHNKKCKEETNVLGQ 408
 QY 399 IPGKMWLVPIYSAVHLDSRY-DQPLNFPWRMOQONNGASSGSGSFTGNNYPMFGG 457
 DB 409 IEKGTNVQVDTWTLTHDPKVMGEDANFERPERME-----SGDELFPYAKG-GYLPFGM 459
 QY 458 GPRLCAGSELAKLEMAVFIHHLVLFKNME 486
 DB 460 GPRICIGMLAMMEKKMLTHILKKYTFE 488

 348 KKEIGSELNMDYKMDFTOCVINEETLGNVYFLH-RKALKDVRKGYDIPSGMKVL 406
 DB 357 DRECPDEPTTFDOLSKLYLECVCYKEALRLPLASVYNNRCKLTTNVLGMEIENGNTIN 416
 QY 407 PVISAVHLDSRY-DQPLNFPWRMOQONNGASSGSGSFTGNNYPMFGGPRLCAGS 465
 DB 417 VDTWTLTHDPKVMGEDANFERPERME-----SGDELFPYAKG-GYLPFGM 467
 QY 466 ELAKLEMAVFIHHLVLFKNME 486

DB 468 RLAMMEMKMLTINIKNTYPE 488

RESULT 15

C912.ARA7H STANDARD: PRT; 500 AA.

AC 065790; Q95Z03; 16-OCT-2001 (Rel. 40, Created)

DT 16-OCT-2001 (Rel. 40, last sequence update)

DT 16-OCT-2001 (Rel. 40, last annotation update)

DE Cytochrome P450 91A2 (EC 1.14.14.1).

GN CYP1A2 OR ATG637430 OR P6G17.80.

OS Arabidopsis thaliana (Mouse-ear cress).

OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae; eurosids II; Brassicales; Brassicaceae; Arabidopsis.

OX NCBI_TaxID=3702;

RN [1]

RP SEQUENCE FROM N.A.

RC STRAIN=cv. Columbia; TISSUE=Seedling;

RX MEDLINE=98281573; PubMed=9620263;

RA Mizutani M., Ward E., Ohta D.:

RT "Cytochrome P450 superfamily in Arabidopsis thaliana: isolation of cDNAs, differential expression, and RFLP mapping of multiple cytochromes P450.";

RL Plant Mol. Biol. 37:39-52(1998).

[2]

RP SEQUENCE FROM N.A.

RC STRAIN=cv. Columbia;

RX MEDLINE=20083488; PubMed=10617198;

RA Mayer K.F.X., Schueller C., Wambutt R., Murphy G., Volckaert G., Pohl T., Duesterhoeft A., Stiekema W., Entian K.-D., Terryn N., Harris B., Ansgore W., Brandt P., Griwell L.A., Rieger M., Weichselgartner M., de Simone V., Obermaier B., Maché R., Mueller M., Kreis M., Delseny M., Pulgomech P., Watson M., Schmidheini T., Reichelt B., Portetelle D., Perez-Alonso M., Boutry M., Bancroft I., Vos P., Homel J., Zimmermann W., Wedler H., Ridley P., Langham S.A., McCullagh B., Billham L., Robben J., Van der Schueren J., Grymoprez B., Chuang Y.-J., Vandenbussche F., Braeken M., Weltjens I., Voet M., Bastiaens I., Aert R., Defoor E., Weltzengerger T., Bothe G., Ramsperger U., Hilbert H., Braun M., Holzer E., Brandt A., Peters S., Van Staveren M., Dirkse W., Moeljan P., Klein Lankhorst R., Rose M., Haut J., Koetler P., Bernieris S., Hempel S., Feldpausch M., Lambers H., Van den Daele H., De Keyser A., Buyschaert C., Gielen J., Villarroel R., De Clercq R., Van Montagu M., Rogers J., Cronin A., Quail M., Bray-Allen S., Clark L., Doggett J., Hall S., Kay M., Lennard N., McIay K., Mayes R., Pettett A., Rajandream M.A., Lyne M., Benes V., Reumann S., Borova D., Bloeker H., Scharfe M., Gilm M., Loehner T.-H., Dose S., de Haan M., Maarse A.C., Schaefer M., Mueller-Auer S., Gabel C., Fuchs M., Fairman B., Granderath K., Dauner D., Herzl A., Neumann S., Agirito A., Vitale D., Liguori R., Piravandi E., Massenot O., Quigley F., Clabaud G., Muendlein A., Felber R., Schrabl S., Hiller R., Schmidt W., Lecharny A., Aubourg S., Chedder F., Cooke R., Berger C., Montfort A., Casacuberta E., Gibbons T., Weber N., Vandenbol M., Barges M., Terol J., Torres A., Perez-Perez A., Purnelle B., Bent E., Johnson S., Tacon D., Jesse T., Heijnen L., Schwarz S., Scholler P., Heber S., Francis P., Bialke C., Frishman D., Haase D., Lemcke K., Mewes H.-W., Stocker S., Zaccaria P., Bevan M., Wilson R.K., de la Bastide M., Habermann K., Parnell L., Dedha N., Gnoj L., Schurz K., Huang E., Spiegel L., Sekon M., Murray J., Sheet P., Cordes M., Abu-Threiden J., Stoneking T., Kalicki J., Graves T., Harmon G., Edwards J., Latelle P., Courtney L., Cloud J., Abbott A., Scott K., Johnson D., Minx P., Bentley D., Fulton B., Miller N., Greco T., Kemp K., Krimer J., Fulton L., Mardis E., Dante M., Pepin K., Hiller L., Nelson J., Spieth J., Ryan E., Andrews S., Geisel C., Layman D., Du H., Ali J., Berghoff A., Jones K., Dione K., Cotton K., Joshi C., Antoniou B., Zidanic M., Strong C., Sun H., Lamar B., Jordan C., Ma P., Zhong J., Preston R., Vill D., Shekher M., Matero A., Shah R., Swaby I.K., O'Shaughnessy A., Rodriguez M., Hoffman J., Tili S., Granat S., Shohdy N., Hasegawa A., Hameed A., Lodi M., Johnson A., Chen E., Marra M., Martienssen R., McCombie W.R.;

RT "Sequence and analysis of chromosome 4 of the plant Arabidopsis thaliana.";

RL Nature 402:769-777(1999).

CC -1- SIMILARITY: BELONGS TO THE CYTOCHROME P450 FAMILY.

CC

CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb.ch/announce/> or send an email to license@isb-sib.ch).

CC

DR EMBL: D78607; BAA28539.1; -

DR EMBL: AL035601; CAB38210.1; -

DR InterPro: IPR001128; Cytochrome_P450.

DR Pfam: PF00067; P450.1.

DR PRINTS: PR00385; P450.

DR PROSITE: PS00086; CYTOCHROME_P450; 1.

KW Oxidoreductase; Monooxygenase; Membrane; Heme.

FT BINDING 438 438 HEME (BY SIMILARITY).

FT CONFLICT 106 106 A -> V (IN REF. 1).

FT CONFLICT 127 127 I -> M (IN REF. 1).

FT CONFLICT 140 140 N -> I (IN REF. 1).

FT CONFLICT 454 454 S -> T (IN REF. 1).

SO SEQUENCE 500 AA; 57555 MW; 0FBA53D207DEA2BA CRC64;

Query Match 11.8%; Score 316; DB 1; Length 500;

Best Local Similarity 25.2%; Pred. No. 5.4e-12;

Matches 135; Conservative 83; Mismatches 199; Indels 118; Gaps 22;

9 LLEPLLLPLSLILLPLILKLR--RNKTRFENLPKSGMPPLGRTIGYKPTATILGD 66

1 MLTYFILP-----LLELVISYKFLYKSTQRFNLPSPSPSPFGVH-LHMKP--PIHR 50

67 FMOOHVSKYKIRYSNLFSEPTIV-----SADAGLRFITONGRLEFECSPSISG 117

51 LIGRYSNGVPIR-SLRFSGRRVYVITSPSLAGEFTGONDVILSSPDLDTAKYVAVNH 109

118 GILGKSMVLVVDMDHDMST-SLNFSLHARLRTI--LIKDYERHTLVYDSMOONSIF 174

110 TVYG---PAPYGDHNRNLRICSOELISSHRLINFQHIKMDILRLTSLRYTQTS-- 163

175 SAODEAKKFF-----NLMAKHITSMDPG-----EEETQOLKEVYTFKGV 216

164 ---NESNDPFIHLEPLSLDLTNNIVRYMTGKRYTGDDVNNKEEALFK----- 211

217 VSAPLNLPGTAYRKALQSRAT-----ILKFERKMEERKLDIKEDQEEFEVKTEDEA 269

212 -----LVYDIAMVSGANHSADYLPILKLGKNFEKRVKAI-----KSMDDI 253

270 EMSKSPHKQRKRDLDLGLVNLKHSNLSTEQIID-----LISLRLRAGHTSSVATLAI 324

254 LQRLIDECRRDKNGMNVNHLISLQOOQPEPYTVIILKIMSMMLAGTSAVTEWAM 313

325 FFLQACPKAWEELREELHETARAKKEGESEL-NMWDYKMDFTQCVINETLRGNVRF 363

314 ANLRMPVEYIEKRSF-----IDKIGKRLIDESDIANVPLQNVNSTFRILFPAPF 367

368 LIPRSPDMDKIGDYVPRDTIVYVAMNAIHROPDEIWEBEKFNPDY---NDGC----- 419

443 GSSTSTGNNT-----MPFGGPRPLCAGSELAKEMAVFIHHLVLRKWEALDQ 492

420 -----GSDYVYVKIMPFNGRRTCPGAGIGQRIIVTLALGSLIQCFEWEVNGKEE 468

Search completed: June 10, 2003, 21:57:52

Job time : 20 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: June 10, 2003, 21:54:48 ; Search time 49 Seconds
(without alignments)
2157.187 Million cell updates/sec

Title: US-09-502-426b-2

Perfect score: 2681

Sequence: 1 MFETEHHTLPLLLPSLLS.....FAFPVDPNGLPIVSRIL 513

Scoring table:

BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 671580 seqs, 206047115 residues

Total number of hits satisfying chosen parameters: 671580

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

SPTREMBL_21:*
1: sp_archaea:*
2: sp_bacteria:*
3: sp_fungi:*
4: sp_human:*
5: sp_invertebrate:*
6: sp_mammal:*
7: sp_mhc:*
8: sp_organelle:*
9: sp_phage:*
10: sp_plant:*
11: sp_rodent:*
12: sp_virus:*
13: sp_vertebrate:*
14: sp_unclassified:*
15: sp_virus:*
16: sp_bacteriap:*
17: sp_archaeap:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	2681	100.0	513	10	064989 arabidopsis
2	2677	99.9	513	10	09SC09
3	951	35.5	474	10	09LKH7
4	866	32.3	512	10	09EX29
5	843.5	31.5	491	10	0941A6
6	822	30.7	478	10	09LNT3
7	815	30.4	382	10	09LY89
8	813.5	30.3	490	10	0941W5
9	789	29.4	464	10	0941C5
10	756.5	28.2	481	10	08M0R4
11	726	27.1	465	10	09FMA5
12	712	26.6	465	10	0940V4
13	711	26.5	465	10	09LH81
14	672.5	25.1	467	10	0949P1
15	658	24.5	463	10	09FH76
16	632	23.6	457	10	065624

17	631	23.5	518	10	09FI38	09fi38 arabidopsis
18	631	23.5	735	10	09IG17	09ig17 arabidopsis
19	622	23.2	443	10	09LJK2	09ljk2 arabidopsis
20	611	22.8	482	10	081077	081077 arabidopsis
21	606	22.6	479	10	08VZC2	08vzc2 arabidopsis
22	597.5	22.3	485	10	08W4T9	08w4t9 taxus cuspi
23	596.5	22.2	497	10	09AXM6	09axm6 taxus cuspi
24	589	22.0	486	10	09L169	09l169 arabidopsis
25	581	21.7	487	10	023384	023384 arabidopsis
26	581	21.7	496	10	09FOV4	09fov4 cucurbita m
27	579.5	21.6	474	10	09SHV7	09shv7 arabidopsis
28	579.5	21.6	485	10	09SJH2	09sjh2 arabidopsis
29	577	21.5	455	10	09LXH8	09lxh8 arabidopsis
30	576.5	21.5	490	10	09C5Y3	09c5y3 arabidopsis
31	575.5	21.5	464	10	004949	004949 arabidopsis
32	569.5	21.2	477	10	09LVT7	09lvt7 arabidopsis
33	547.5	20.4	489	10	09ZV72	09zv72 arabidopsis
34	546.5	20.4	444	10	08S917	08s917 arabidopsis
35	546.5	20.4	489	10	09C5Y2	09c5y2 arabidopsis
36	518	19.3	504	10	09SNG3	09sng3 oryza sativ
37	502	18.7	499	10	09AXH9	09axh9 hordeum vul
38	497.5	18.6	460	10	09STN2	09stn2 arabidopsis
39	453	16.9	497	11	08VIL0	08vil0 ratius norv
40	435	16.2	224	10	09SDM6	09sdm6 helianthus
41	421	15.7	394	10	08W1S6	08w1s6 zea mays su
42	414	15.4	394	10	08W1T1	08w1t1 zea mays su
43	414	15.4	394	10	08W1S9	08w1s9 zea mays su
44	413	15.4	394	10	08W1J5	08w1j5 zea mays su
45	412	15.4	394	10	08W1V5	08w1v5 zea mays su

ALIGNMENTS

RESULT 1
ID 064989 PRELIMINARY; PRT; 513 AA.
AC 064989:
DT 01-AUG-1998 (TREMBLrel. 07, Created)
DT 01-AUG-1998 (TREMBLrel. 07, Last sequence update)
DT 01-MAR-2002 (TREMBLrel. 20, Last annotation update)
DE Steroid 22-alpha-hydroxylase.
GN DWF4.
OS Arabidopsis thaliana (Mouse-ear cress).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;
OC eurosids II; Brassicales; Brassicaceae; Arabidopsis.
OX NCBI_TaxID=3702;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=MS-2;
RX MEDLINE=98158690; PubMed=9490746;
RA Choe S., Dalkes B.P., Fujlaka S., Takatsuto S., Sakurai A.,
RA Feldmann K.A.;
RT "The DWF4 gene of Arabidopsis encodes a cytochrome P450 that mediates
RT multiple 22alpha-hydroxylation steps in brassinosteroid
RT biosynthesis."
RL Plant Cell 10:231-243(1998).
RC -1- SIMILARITY: BELONGS TO THE CYTOCHROME P450 FAMILY.
DR EMBL: AF044216; AAC05093.1; -
DR InterPro: IPR001128; Cytochrome_P450.
DR Pfam: PF00067; P450: 1.
DR PROSITE: PS00086; CYTOCHROME_P450; UNKNOWN_1.
KW Heme; Monooxygenase; Oxidoreductase.
SQ SQUONCE 513 AA: 58867 MW; B1639BD9A4DA6F3 CRC64;

Query Match 100.0%; Score 2681; DB 10; Length 513;
Best Local Similarity 100.0%; Pred. No. 1.6e-184;
Matches 513; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MFETEHHTLPLLLPSLLSLFLILKRNKRTRFNNLPKSGMPFLETIGYLRPYT 60
DB 1 MFETEHHTLPLLLPSLLSLFLILKRNKRTRFNNLPKSGMPFLETIGYLRPYT 60

QY 61 ATTGADPMOQHVSKYKTYRNSLFGPEPTIVSADAGLNFILQNEGRLEFCSYPRISGIGIL 120
 DB 61 ATTGADPMOQHVSKYKTYRNSLFGPEPTIVSADAGLNFILQNEGRLEFCSYPRISGIGIL 120
 QY 121 GKMSMLVGVGMHMRMSISLNFISHARLRTILKDVDRHTLFLVDSMOQNSIFSADQEA 180
 DB 121 GKMSMLVGVGMHMRMSISLNFISHARLRTILKDVDRHTLFLVDSMOQNSIFSADQEA 180
 QY 181 KKFENLMAKHIMSMDPEEETEDOLKKEYVTFMKGVSAPIPLPGTAYHAKLQSRATILK 240
 DB 181 KKFENLMAKHIMSMDPEEETEDOLKKEYVTFMKGVSAPIPLPGTAYHAKLQSRATILK 240
 QY 241 FIERKMEERKIDIEEDQEEVEVTEDEAEMSKSDHVRKORTDDLLGVLKHSNLSTEQ 300
 DB 241 FIERKMEERKIDIEEDQEEVEVTEDEAEMSKSDHVRKORTDDLLGVLKHSNLSTEQ 300
 QY 301 ILDLILSLFAGHETSSVAIALAIFPLQACPAVEELREHELEIARAKKELESELNMD 360
 DB 301 ILDLILSLFAGHETSSVAIALAIFPLQACPAVEELREHELEIARAKKELESELNMD 360
 QY 361 YKKMDFQCYINETLRGLNVVRLHRRKALDVRYGYDIPSGMKVLPVISAHLDSRYD 420
 DB 361 YKKMDFQCYINETLRGLNVVRLHRRKALDVRYGYDIPSGMKVLPVISAHLDSRYD 420
 QY 421 QPNLFNPMWQOQNNGASSSGSSTWGNMYPGGGPRLCAGSELAKLEMAVFIHILV 480
 DB 421 QPNLFNPMWQOQNNGASSSGSSTWGNMYPGGGPRLCAGSELAKLEMAVFIHILV 480
 QY 481 LKFNNELAEDDQPPAFPFVDPNGLPPIRVSRIL 513
 DB 481 LKFNNELAEDDQPPAFPFVDPNGLPPIRVSRIL 513

RESULT 2

Q9SC09 PRELIMINARY: PRT: 513 AA.
 AC Q9SC09
 DT 01-MAY-2000 (TREMBlrel. 13, Created)
 DT 01-MAY-2000 (TREMBlrel. 13, Last sequence update)
 DT 01-JUN-2002 (TREMBlrel. 21, Last annotation update)
 DE Steroid 22-alpha-hydroxylase (DMF4) (At3g50660/T3A5_40).
 GN T3A5_40.
 OS Arabidopsis thaliana (Mouse-ear cress).
 OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;
 OC eurosids II; Brassicales; Brassicaceae; Arabidopsids.
 OC NCBI_Taxid=3702;
 RN [1]
 RP SEQUENCE FROM N.A.
 RA Bloeker H., Meves H.W., Lemcke K., Mayer K.F.X., Quettler F.,
 RA Salanoubat M.;
 RL Submitted (NOV-1999) to the EMBL/GenBank/DBJ databases.
 RN [2]
 RP SEQUENCE FROM N.A.
 RA EU Arabidopsis sequencing project;
 RL Submitted (JAN-2000) to the EMBL/GenBank/DBJ databases.
 RN [3]
 RP SEQUENCE FROM N.A.
 RA Cheuk R., Chen H., Kim C.J., Koesema E., Meyers M.C., Banh J.,
 RA Bowser L., Carninci P., Dale J.M., Goldsmith A.D., Hayashizaki Y.,
 RA Ishida J., Jiang P.X., Jones T., Kamliya A., Karlin-Neumann G.,
 RA Kawai J., Lam B., Lee J.M., Lin J., Liu S.X., Miranda M., Narusaka M.,
 RA Nguyen M., Onodera C.S., Palm C.J., Pham P.K., Quach H.L., Sakurai T.,
 RA Satou M., Seki M., Southwick A., Tang C.C., Toriumi M., Yamada K.,
 RA Yamamura Y., Yu G., Yu S., Shinozaki K., Davis R.W., Theologis A.,
 RA Ecker J.R.;
 RT "Arabidopsis cDNA clones."
 RL Submitted (AUG-2001) to the EMBL/GenBank/DBJ databases.
 RN [4]
 RP SEQUENCE FROM N.A.
 RA Cheuk R., Chen H., Kim C.J., Meyers M.C., Shin P., Banh J.,
 RA Bowser L., Carninci P., Chang E., Dale J.M., Goldsmith A.D.,

RA Hayashizaki Y., Ishida J., Jones T., Kamliya A., Karlin-Neumann G.,
 RA Kawai J., Lam B., Lee J.M., Lin J., Miranda M., Narusaka M.,
 RA Nguyen M., Onodera C.S., Palm C.J., Pham P.K., Quach H.L., Sakurai T.,
 RA Seki M., Southwick A., Tang C.C., Toriumi M., Wu H.C., Yamada K.,
 RA Yamamura Y., Yu G., Yu S., Shinozaki K., Davis R.W., Theologis A.,
 RA Ecker J.R.;
 RT "Arabidopsis ORF clones."
 RL Submitted (MAR-2002) to the EMBL/GenBank/DBJ databases.
 CC -1- SIMILARITY: BELONGS TO THE CYTOCHROME P450 FAMILY.
 DR EMBL: AL132979; CAB62435.1; -;
 DR EMBL: AF412114; AAL06567.1; -;
 DR EMBL: AY090266; AAL09027.1; -;
 DR InterPro: IPR001128; Cytochrome_P450.
 DR Pfam: PF00067; P450; 1.
 DR PRINTS: PR00385; P450.
 DR PROSITE: PS00086; CYTOCHROME_P450; UNKNOWN_1.
 KM Heme: Monooxygenase; Oxidoreductase.
 SQ SEQUENCE 513 AA: 58867 MW: 81639BD9A5D7C93 CRC64;

Query Match 99.9%; Score 2677; DB 10; Length 513;
 Best Local Similarity 99.8%; Pred. No. 3e-184;
 Matches 512; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 1 MEETHHTLPLLLPILSLILFLILKRRRKTRFNLPPKSGMPFGITGYLKPPT 60
 DB 1 MEETHHTLPLLLPILSLILFLILKRRRKTRFNLPPKSGMPFGITGYLKPPT 60
 QY 61 ATTGADPMOQHVSKYKTYRNSLFGPEPTIVSADAGLNFILQNEGRLEFCSYPRISGIGIL 120
 DB 61 ATTGADPMOQHVSKYKTYRNSLFGPEPTIVSADAGLNFILQNEGRLEFCSYPRISGIGIL 120
 QY 121 GKMSMLVGVGMHMRMSISLNFISHARLRTILKDVDRHTLFLVDSMOQNSIFSADQEA 180
 DB 121 GKMSMLVGVGMHMRMSISLNFISHARLRTILKDVDRHTLFLVDSMOQNSIFSADQEA 180
 QY 181 KKFENLMAKHIMSMDPEEETEDOLKKEYVTFMKGVSAPIPLPGTAYHAKLQSRATILK 240
 DB 181 KKFENLMAKHIMSMDPEEETEDOLKKEYVTFMKGVSAPIPLPGTAYHAKLQSRATILK 240
 QY 241 FIERKMEERKIDIEEDQEEVEVTEDEAEMSKSDHVRKORTDDLLGVLKHSNLSTEQ 300
 DB 241 FIERKMEERKIDIEEDQEEVEVTEDEAEMSKSDHVRKORTDDLLGVLKHSNLSTEQ 300
 QY 301 ILDLILSLFAGHETSSVAIALAIFPLQACPAVEELREHELEIARAKKELESELNMD 360
 DB 301 ILDLILSLFAGHETSSVAIALAIFPLQACPAVEELREHELEIARAKKELESELNMD 360
 QY 361 YKKMDFQCYINETLRGLNVVRLHRRKALDVRYGYDIPSGMKVLPVISAHLDSRYD 420
 DB 361 YKKMDFQCYINETLRGLNVVRLHRRKALDVRYGYDIPSGMKVLPVISAHLDSRYD 420
 QY 421 QPNLFNPMWQOQNNGASSSGSSTWGNMYPGGGPRLCAGSELAKLEMAVFIHILV 480
 DB 421 QPNLFNPMWQOQNNGASSSGSSTWGNMYPGGGPRLCAGSELAKLEMAVFIHILV 480
 QY 481 LKFNNELAEDDQPPAFPFVDPNGLPPIRVSRIL 513
 DB 481 LKFNNELAEDDQPPAFPFVDPNGLPPIRVSRIL 513

RESULT 3

Q9LKH7 PRELIMINARY: PRT: 474 AA.
 AC Q9LKH7
 DT 01-OCT-2000 (TREMBlrel. 15, Created)
 DT 01-OCT-2000 (TREMBlrel. 15, Last sequence update)
 DT 01-JUN-2002 (TREMBlrel. 21, Last annotation update)
 DE Cytochrome P450.
 GN CYP90A2.
 OS Vigna radiata.
 OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;
 OC eurosids II; Fabales; Fabaceae; Papilionoideae; Phaseoleae; Vigna.

RP SEQUENCE FROM N.A.
RC STRAIN-CV. COLUMBIA:
RA Lin X., Kaul S., Town C.D., Benito M., Creasy T.H., Haas B.J., Wu D.,
RA Maiti R., Ronning C.M., Koo H., Fujii C.Y., Utterback T.R.,
RA Barnstead M.E., Bowman C.L., White O., Niernan W.C., Fraser C.M.,
RT "Arabidopsis thaliana chromosome 1 BAC T9124 genomic sequence.",
RL Submitted (NOV-2000) to the EMBL/GenBank/DBD databases.
CC - SIMILARITY: BELONGS TO THE CYTOCHROME P450 FAMILY.
DR EMBL: AC013966; AAC30983.1; -
DR InterPro: IPR001128; CYTOCHROME_P450.
DR Pfam: PF00067; p450; 1.
DR PRINTS: PRO0385; P450.
DR PROSITE: PS00086; CYTOCHROME_P450; UNKNOWN_1.
KW Heme: Monooxygenase; Oxidoreductase.
SQ SEQUENCE 512 AA; 58297 MW; 15900C034BF319FA CRC64;

Best Local Similarity 33.8%; Pred. No. 4.2e-54;
Matches 181; Conservative 105; Mismatches 173; Indels 76; Gaps 7

```

Db      11  LLSVSSSTFFIAFIITIFLLAGIARRKKRAPHPLPGSGMWGLJIDITFAMLNAAVAGSPSS  70
Oy      67  FMOQHVSK-----YKTIYSMLFGEPIITVSADACLANFFIL  101
Db      71  FVEKQIKTKFEVSLICSVLLILIKRPNSGFNEIRYGRITSCSLFKRMVAVSADPFPNFRIM  130
Oy      102  QNEGRLEECSTPYRISIGLIGKSMVLVGMIDMRDRSTSLNFIASHARLTLLDQVRHT  161
Db      131  QNNGKLFQSSYRPSFDFLDVGKGVITVHDDQORLHSTIASSMRRHDQIKTEFLEVIYVM  190
Oy      162  LEVLDSMOQNSIFSAODEAKKFTFENIMAKHIMSDPGEETEOUKKEVTFPMKVSAPL  221
Db      191  LQPLSNFKKEGEVLLIDIDICKRVAIHLMWQLIGVS-SESEVDENESOLFSDVDCSLVPI  249
Oy      222  NLEPGTAYHKALOSRAITLKFIERKKEERKLDIKEEDQEEBEVTEDEAEEMSKSDHVAKOR  281
Db      250  DLFGFTYNNKRAKKRIITIKKTIETKRLQNKAAAD-----T  286
Oy      282  TDDLLIGWLKSHNSLSTEOILDLILSLFAGHETISSVAIALAIFLQACPRAVELEEBH  341
Db      287  AGNGVIGRLLEEESLFPNESMADFIINLLFAGNETTSKTMFLAVYFLTHCPKAMQOLDEEH  346
Oy      342  LELARRAKKELGSESLWDDYKKMDPTOCYINTELTGLVNVFIRLRAKLDKVRYGCIYPS  401
Db      347  -----DRLAGSLMTWQDKTMDTFOCYIDETLGLGIAIWLKREAKEDVSYDIYIPK  399
Oy      402  GMRVLVPIASVHLNDSRYDQPILEFNPWR-----QQQNSAGSSSGSGSFSTWGNMYMFG  456
Db      400  GCEVYEPFLSVHLEDESZYKESLSPNWRMLDEPQOKRNMTSP-----FYCPFG  449
Oy      457  GGRILCAGSLAKLEMAVFIIHILVILKFNNELAEDDQAPAFVDPNGLPIRYAR  511
Db      450  GGRFRCGAEIARLQIALFLHYFTIYYKWTOLKEDRISFPSPSARLVNMGFKIQLNR  504

```

0941A6		
ID	0941A6	PRELIMINARY; PRT: 491 AA.
AC	0941A6;	
DT	01-DEC-2001 (TREMBLrel. 19, Created)	
DT	01-DEC-2001 (TREMBLrel. 19, Last sequence update)	
DT	01-JUN-2002 (TREMBLrel. 21, Last annotation update)	
DE	CYP9D0.	
CN	CYP9D0.	
OS	Aradidopsis thaliana (Mouse-ear cress).	
OC	Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;	
OC	Spermatophyta; Magnoliophyta; eudicotyledons: core eudicots; Rosidae	
OC	eumastids II; Brassicales; Brassicaceae; Arabidopsis.	
OX	NCBI_TaxID=3702;	
NP	[1]	
RP	SEQUENCE FROM N.A.	

RC STRAIN=CV, COLUMBIA;
RA Shimada Y.;
RT "P450 gene repressed by brassinosteroid."; [PubMed](#)
RL submitted (JUL-2001) to the EMBL/Genbank/DBJ databases.
CC -1'SIMILARITY: BELONGS TO THE CYTOCHROME P450 FAMILY.
DR EMBL; AB066286; BAB62109.1; "-"
DR InterPro: IPR001128; Cytochrome_P450.
DR InterPro: IPR000504; RNA_rec_mot.
DR Pfam; PF00067; P450; 1.
DR PROSITE; PS00086; CYTOCHROME_P450; UNKNOWN_1.
DR PROSITE; PS00030; RRM_RNP_1; UNKNOWN_1.
KW Heme; Monooxygenase; Oxidoreductase
SQ SEQUENCE 491 AA; 56153 MW; 02PFB908A81995A40 CRC64;

RP SEQUENCE FROM N.A.
 RA Arabidopsis sequencing project;
 RL Submitted (APR-2000) to the EMBL/GenBank/DBJ databases.
 CC -1- SIMILARITY: BELONGS TO THE CYTOCHROME P450 FAMILY.
 DR EMBL: A1163817; CAB87779.1; -
 DR InterPro: IPR001128; Cytochrome_P450.
 DR Pfam: PF00067; P450; 1.
 DR PROSITE: PS00086; CYTOCHROME_P450; UNKNOWN_1.
 KW Heme; Hypothetical protein; Monooxygenase; Oxidoreductase.
 SQ SEQUENCE 382 AA; 43889 MW; 1BC5685AB73BAE30 CRC64;

Query Match 30.4%; Score 815; DB 10; Length 382;
 Best Local Similarity 34.2%; Pred. No. 1.3e-50;
 Matches 159; Conservative 87; Mismatches 127; Indels 92; Gaps 4;

```

QY 45 GMPFGETISGLKPYATATIDGFMGOHSHKYSKIRSNLFGPTVSADAGINRITONE 104
DB 2 GMPFGETISFPKPHRSISIGTFLQOVRSRGKVKSNICGGKAVVSCDOEIMNFIQNE 61
QY 105 GRLFECSPRSTIGILGKMSMLVLGDMHRDMSISLNFSLSHARLTILKDVREHTLFY 164
DB 62 GLFTSDYPRKAMHMDILGKYSLLATGEIHRKLNKIVISFINLTKSPDLCAENLSTI 121
QY 165 LDMOONSIFSAODEAKKFTFNLMKHTMSMDPGEETEDQKKKEYTFMGVSAPIMLP 224
DB 122 LKSWNCRVEVEFHKEVKFTLSVMYNQLSTKPEDPARLVYLODFLSYMKGFISLPIPLP 181
QY 225 STAHKALQSRATITLTERKMEERKLDIKEEDQEEVEKTEDEAKMSKSDHVKORTDD 284
DB 182 GTGTNAIKVRSNNIHNQNIIEEDMNNAIREDLDLSTISMD----- 224
QY 285 DLGVLKHSNLSLSTEQILDLISLFGHETSSVAIALAFPLQACPKAVEELREHLEI 344
DB 225 -----EEHAAI 230
QY 345 ARAKKEGSELNMDYRKMDPTQCVINETRLGNNVRLHRRALKDVRKGYDIPBGK 404
DB 231 -RAKKGDEL-LNNEEDQKMEFTQCVSEALRCGNIKYTHRAKTHDINKNEYVIRPGWK 288
QY 405 VLPYISAVHLDNSYDOPNLFNPRWQOONNGASSSGSFSFGNNYMPFGGPRLCAG 464
DB 289 VFPTFAVHLDPSSLHEPFEFNPWRMTKT-----AFGGGVAVCPG 329
QY 465 SELAKLEMAVFIHHLVLFKNMELAEEDOPAPFPVDPENGLPIRV 509
DB 330 GEIGKIQIAFFLHVLVSTYRKIKSDMPIAHFYVERKGMLEI 374

```

RESULT 8

ID 0941W5 PRELIMINARY; PRT; 490 AA.
 AC 0941W5;
 DT 01-DEC-2001 (Tremblrel. 19, Created)
 DT 01-DEC-2001 (Tremblrel. 19, Last sequence update)
 DT 01-JUN-2002 (Tremblrel. 21, Last annotation update)
 DE Cytochrome P450-like protein.
 GN P0419B01.5.
 OS Oryza sativa (Rice).
 OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 OC Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
 OC NCBI_TaxID=4530; Oryza.
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=CV. NIPPONBARE;
 RA Sasaki T., Matsumoto T., Yamamoto K.;
 RT "Oryza sativa nipponbare(GA3) genomic DNA, chromosome 1, PAC
 clone:p0419B01.";
 RL Submitted (FEB-2001) to the EMBL/GenBank/DBJ databases.
 CC -1- SIMILARITY: BELONGS TO THE CYTOCHROME P450 FAMILY.
 DR EMBL: AP00244; BAB56089.1; -
 DR InterPro: IPR001128; Cytochrome_P450.
 DR Pfam: PF00067; P450; 1.

DR PROSITE: PS00086; CYTOCHROME_P450; UNKNOWN_1.
 KW Heme; Monooxygenase; Oxidoreductase.
 SQ SEQUENCE 490 AA; 54824 MW; 9BC2853BBAFBB88P CRC64;

Query Match 30.3%; Score 813.5; DB 10; Length 490;
 Best Local Similarity 34.2%; Pred. No. 2.4e-50;
 Matches 173; Conservative 106; Mismatches 136; Indels 71; Gaps 9;

```

QY 27 LTKRRNR-----KTRFNLPKSGKMPFGETIGLYKPYATATIDGFMGOHSHKYSK 77
DB 28 LLRRRRRAGSGKDAAAARLPSPSGFMPVGETLFEVSOASPREAVDRKRIHGS 87
QY 78 -IYRSNLGEPETIYASDAGINRFTLQNEGRLFECSTYRSIGILGKMSMLVLGDMHRD 136
DB 88 AVFSHLPFGSATVTTAAAEVSRFVQSDARAFVFWYRSLTELMGSSILLINGALQRRV 147
QY 137 RSISLNLSHARLTILKDVREHTLFVLDMSOONSIFSAODEAKKFTFNLMKHTMSMD 196
DB 148 HGLVGAFFKSSHLKSQTLADMRRRLSPALSPDSSLHQLAKSVPEILVGLIGLE 207
QY 197 PGEETEDQKKKEYTFMGVSAPIMLPGTAHKAQSRATITLTERKMEERK----- 250
DB 208 AG-DEMQLQOQOEFTVGLMSLPKLPGRILYRSLOAKKARLQRIIRERARRAAA 266
QY 251 -----LDIKEEDQEEVEKTEDEAKMSKSDHVKORTDDLGLVLSNLSLSTEQIIL 304
DB 267 SPPRDAIDVLIGDSD-----LDELISDN 292
QY 305 ILSLFAGHETSSVAIALAFPLQACPKAVEELREHLEIARAKKEGSELNMDYRK 364
DB 293 MIDLMIPAEDEVVLTTLVAKFLSECPDLHLHOLEENIOLKRRKTDMEGT-LQMDYMSL 351
QY 365 DFTQCVINETRLGNNVRLHRRALKDVRKGYDIPSGKVPVYSAVHLNSRYDOPNL 424
DB 352 SFQOHVITETRLGNIIGIMRAVDYEVKHLIPKGCYVYRSHLDDTLDEPK 411
QY 425 FNPWRWQOONNGASSSGSFSFGNNYMPFGGPRLCAGSELAKLEMAVFIHHLVLFKN 484
DB 412 FNPWRKED-----MSNSST-----PFGGGRQLPGDLRLAKSIFLHLVTSFR 459
QY 485 WELAEEDOPAPFPVDPENGLPIRV 510
DB 460 W-VAEDHIVNPTVIRKGMPIRV 484

```

RESULT 9

ID 091IC5 PRELIMINARY; PRT; 464 AA.
 AC 091IC5;
 DT 01-OCT-2000 (Tremblrel. 15, Created)
 DT 01-OCT-2000 (Tremblrel. 15, Last sequence update)
 DT 01-JUN-2002 (Tremblrel. 21, Last annotation update)
 DE Cytochrome P450-like protein.
 OS Arabidopsis thaliana (Mouse-ear cress).
 OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;
 OC euroids II; Brassicales; Brassicaceae; Arabidopses.
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=COLUMBIA;
 RA Kaneko T., Kato T., Sato S., Nakamura Y., Asamizu E., Tabata S.;
 RL Submitted (MAR-2000) to the EMBL/GenBank/DBJ databases.
 RN [2]
 RP SEQUENCE FROM N.A.
 RC STRAIN=COLUMBIA;
 RX MEDLINE=20363099; Pubmed=10907853;
 RA Nakamura Y.;
 RT "Structural analysis of Arabidopsis thaliana chromosome 3. II.
 TAC and BAC clones.";
 RL DNA Res. 7:217-221(2000).
 CC -1- SIMILARITY: BELONGS TO THE CYTOCHROME P450 FAMILY.

DR EMBL: AF001307; BAB01922.1; -
 DR InterPro: IPR001128; Cytochrome_P450.
 DR InterPro: IPR000504; RNA_rec_mot.
 DR Pfam: PF00067; P450; 1.
 DR PRINTS: PR00385; P450.
 DR PROSITE: PS00086; CYTOCHROME_P450; UNKNOWN_1.
 DR PROSITE: PS00030; RRM_RNP_1; UNKNOWN_1.
 DR Heme; Monooxygenase; Oxidoreductase.
 KW SEQUENCE 464 AA; 52864 MW; A65E094665E51908 CRC64;
 SQ
 Query Match 29.4%; Score 789; DB 10; Length 464;
 Best Local Similarity 36.5%; Pred. No. 1.3e-48;
 Matches 163; Conservative 101; Mismatches 145; Indels 38; Gaps 8;
 QY 40 PPKSGMPFLGETIGYLRKYATATLIGDMQGVSKRYKRNLSNGEPTIYSADAGLRF 99
 DB 52 PHGSLGMPVIGETIEFVSAYSDRSPDMKRLMYGVFKSHFGTATITSDAEVRA 111
 QY 100 ILNEGRLFECSYRISIGILGKWSMLVYGDHMRMSISLNLSHARLTILLKDYER 159
 DB 112 VLQSDSTAFVPEFYKTYRELKSGSILLINGSLRRRFGLVGSFLKSPILKAQIVRDMK 171
 QY 160 HTLEVLDSMOONSIFSAODEAKKFTFNLMAKHMSMDGEEETQLKKEYVTGPKGVYA 219
 DB 172 FLSESMDSDDQPVLLQDVSKTYAFVLAALISVEKG-RDLELKEFEFNISGLMSL 230
 QY 220 PLNLPGTAYHKAQSRAATILKFERKMERKLDIKEEDQEEVEYKTEDEAM---SKSDH 276
 DB 231 PINPGTQLHRSLOAKKNMYKQVERIEGKIR--KTNKKEEDVIANKVYVVLKDSSEH 288
 QY 277 VRKQRTDLDLGVNLKSNSTEOILDLISLFGHETTSVAIALAIFLOAPKAAVEE 336
 DB 289 -----LTH-NLIANNMIDMI---PGHDSVPVLLITLAKVFSDSPPAALNL 329
 QY 337 LREHLEIARAKKELGSELNMDYKHMDFQCYINETLRIGNVYRFLHRLKADVRKYG 396
 DB 330 LTERNMKL-KSLKELTEBPYLYNDYLSLPFOKVTITELRGNGNYITGWRAMKADVEIKG 388
 QY 397 YDIPSGKVLVPIVAHLDSRYDQPNLFNFRWQOONNGASSGSGSFSTWGNMYMFG 456
 DB 389 YVLPKGCPLAYLHSLVHLDKLYESPYPKFNFMWQERDMNTSS-----FSFPG 436
 QY 457 GGRLCAGSELAKLEMAVFIHHVLYKF 483
 DB 437 GGORLCPGLDLARLETSVFLHHLVTRF 463
 RESULT 10
 OSMORA PRELIMINARY; PRT; 481 AA.
 AC OSMORA;
 DT 01-MAR-2002 (TREMBlrel. 20, Created)
 DT 01-MAR-2002 (TREMBlrel. 20, Last sequence update)
 DT 01-JUN-2002 (TREMBlrel. 21, Last annotation update)
 DE Putative cytochrome P450-like protein.
 GN SB32H7.4.
 OS Sorghum bicolor (Sorghum) (Sorghum vulgare).
 OC Eukaryota; Viridiplantae; Streptophyta; Tracheophyta;
 OC Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACC clade;
 OC Panicoideae; Andropogoneae; Sorghum.
 OX NCBI_Taxid=4558;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=CV. BTX623;
 RA Park Y.-J., Ramakrishna W., Sanmiquel P., Emberton J., Bennetzen J.;
 RL Submitted (JAN-2002) to the EMBL/GenBank/DBJ databases.
 RN [2]
 RP SEQUENCE FROM N.A.
 RC STRAIN=CV. BTX623;
 RA Liaca V., Young S., Koychok S., Messing J.;
 RL Submitted (JAN-2002) to the EMBL/GenBank/DBJ databases.
 DR EMBL: AF466201; AAL73972.1; -
 DR InterPro: IPR001128; Cytochrome_P450.

DR Pfam: PF00067; P450; 1.
 DR PROSITE: PS00086; CYTOCHROME_P450; UNKNOWN_1.
 DR SEQUENCE 481 AA; 54984 MW; 6145CFEFCOFD7AFE CRC64;
 SQ
 Query Match 28.2%; Score 756.5; DB 10; Length 481;
 Best Local Similarity 33.5%; Pred. No. 2.9e-46;
 Matches 170; Conservative 104; Mismatches 190; Indels 43; Gaps 7;
 QY 15 LPSSLISLFLILK-----RRNRTRNLPPKSGMPFLGETIGYLRKYATATTLG 65
 DB 6 LAALSVTLIGALLLMAFWMNMYGRGCGEGLLPFGSLTEFFPASPFLLELV 65
 QY 66 DFMQGVSKRYKRYNSNGEPTIYSADAGLRFILQNGRLPECSYPSISIGILGKWSH 125
 DB 66 PEFKRRLERFGDIFPTNIVGEMIVSLDELARVLOQDERGFQIYPPSFMRLIGADNM 125
 QY 126 LVLYGDMHRDMSISLNLSHARLTILLKDYERHTLFLVDSMQONSIFSADQEAKEKTF 185
 DB 126 VSMGLPRLHRIINLRLFGPEALRLVLRDVORSARDELRSMLDRPEVEVKTATSRMTF 185
 QY 186 NLMAKHMSMDGEEETQLKKEYVTGPKGVYAPNLPGTAYHKAQSRAATILKFERK 245
 DB 186 GYTAKLKLSHD-DVASSGSLMKCFDAWTKGMSFPCYVGTAIFYRCMGQRKNMYKVKQ 244
 QY 246 MEERKLDIKEEDQEEVEYKTEDEAEKSDHVRKQRTDLDLGVNLKSNLSTEOI-LDI 304
 DB 245 LDERR-----NGAERKTVDFDL-----VIDELDKPMSIMSESIALNL 282
 QY 305 ILSLFGHETTSVAIALAIFLOAPKAAVEELREHLEIARAKKELGSELNMDYKYM 364
 DB 283 LFLLFASHETTSMLGTYLTKLFTLNPISLOLTELTEHEKIMERRD-PSDITWEYKSM 341
 QY 365 DFTQCVINETLRIGNVYRFLHRLKADVRKGYDIPSGKVLVPIVAHLDSRYDQPNL 424
 DB 342 KFTSHVHESLRLAIAIPVYRQANDVHIKETTIPESKIMICPSAALINSVYDPLA 401
 QY 425 FNPWMOQONNGASSGSGSFSTWGNMYMFGGGRPLCAGSELAKLEMAVFIHHVLYKFN 484
 DB 402 FNPWMKDTPPEVGS-----KDFMAFGGLRLCGAEGFAKLQAMFLHYLVTNFR 452
 QY 485 WELAEDDQPEAFPEFDPFGULPIRYSR 511
 DB 453 WKALSKGTMTLPGRLPFGPHILOLHK 479
 RESULT 11
 O9FMA5 PRELIMINARY; PRT; 465 AA.
 AC O9FMA5;
 DT 01-MAR-2001 (TREMBlrel. 16, Created)
 DT 01-MAR-2001 (TREMBlrel. 16, Last sequence update)
 DT 01-MAR-2002 (TREMBlrel. 20, Last annotation update)
 DE Cytochrome P450 (Brassinosteroid-6-oxidase).
 GN BR60X1.
 OS Arabidopsis thaliana (Mouse-ear cress).
 OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;
 OC eustos II; Brassicales; Brassicaceae; Arabidopsids.
 OX NCBI_Taxid=3702;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=COLUMBIA;
 RC MEDLINE=98290546; PubMed=9628582;
 RA Sato S., Kaneko T., Kotani H., Nakamura Y., Asamizu E., Miyajima N.,
 RA Tabata S.;
 RT "Structural analysis of Arabidopsis thaliana chromosome 5. IV.
 RT Sequence features of the regions of 1,456,315 bp covered by nineteen
 RT physically assigned PI and TAC clones.";
 RL DNA Res. 5:41-54(1998).
 RN [2]
 RP SEQUENCE FROM N.A.
 RC STRAIN=COLUMBIA;
 RC MEDLINE=21295570; PubMed=11402205;

RA Shimada Y., Fujioaka S., Miyauchi N., Kushihiro M., Takatsuto S.,
 RA Nomura T., Yokota T., Kamiya Y., Bishop G., Yoshida S.,
 RT "Brassinosteroid-6-oxidases from Arabidopsis and tomato catalyze
 RT multiple C-6 oxidations in brassinosteroid biosynthesis.";
 RL Plant Physiol. 126:770-779 (2001).
 CC -1- SIMILARITY: BELONGS TO THE CYTOCHROME P450 FAMILY.
 DR EMBL: AB009048; BAB08653.1; -
 DR EMBL: AB033868; BAB0858.1; -
 DR InterPro: IPR001128; Cytochrome_P450.
 DR Pfam: PF00067; P450; 1.
 DR PRINTS: PR00385; P450.
 DR PROSITE: PS00086; CYTOCHROME_P450; UNKNOWN_1.
 DR Heme; Monooxygenase; Oxidoreductase.
 SO SEQUENCE 465 AA; 53767 MW; 0C0D459C9C866D1F CRC64;

Query Match 27.1%; Score 726; DB 10; Length 465;
 Best Local Similarity 32.3%; Pred. No. 4, 3e-44;
 Matches 166; Conservative 91; Mismatches 181; Indels 76; Gaps 10;

QY 12 LLLPSSLILFLILKRRN--KTRFNLPPGKSGWPLGTTGYLKPATATLGDPMQ 69
 DB 11 LLIIVSLCSALL-----RMQMRKTKNGLPPTGMPFIFGTEFLKQGP-----NFMK 59
 QY 70 OHVSKYKIRSNLFGEPPIVSADAGLNRFILONGRLFECSYPRISIGILGKSMVLV 129
 DB 60 NORLRYGSFFKSHLIGCPITLMSQEVKRIILKSKGLVPGYQSMIDLIGTCNMAVH 119
 QY 130 GDMHDMRSISLNFSLHARLRTILIKDVERHTLVDSMOONSIFSAODEAKKFTF--- 185
 DB 120 GSHHLMKSLSLISLSTMDHILPKVDHFMRSYLDQNMNLEVIDIDQKHKHAFSL 179
 QY 186 -----NLMAKHMSMDPEEETEOLKEEYTFMKGYSAPLNLFGLYHKAQSRATILK 240
 DB 180 TQIAGNLKRPV-----EEKTAFLVGLTSPDIDPGTNYRCQIQAANNIDR 229
 QY 241 FIEKMERKIDIEEEOEEVEKTEDEAEMSKSDHYRKORTDDLLGWLKHSN-----L 236
 DB 230 LIRLMOERR-----DSGE-----TTDMGLYLMKKEGNRYPL 262
 QY 297 STEQILDLILSLFAGHETSSVAIALAIFLOCPKAVEELREHLEIARAKKEGSEL 356
 DB 263 TDEIRDOVYVTLISGTEVSTISMAKLYLHDHPALQELRAEHLAFREKRQ--DEPL 320
 QY 357 NMDYKMDTQCVINETLRLGNVRFILRRKALDKVRYKGYDIPSGKVLPIVSAVHLN 416
 DB 321 GLEDVSMKFTRAVIYETSLRATIVNGYLKTRDLEINGYLIPKMRIVYTRREINYDA 380
 QY 417 SRYDQNLFPFMRMOONNGASSSGSSTWGNMYMPFGGPRLCAGSELAKLENAVPI 476
 DB 381 NLEDPLIFPMRMKKSLESQ-----NSCFVGGGTGLCPGKELGIVEISSFL 429
 QY 477 HHVLYKFWELAEDDQPFAPFVDFPGLPIRVS 510
 DB 430 HFFVTRYWEELIGDELWVFPVFAKGFHLRIS 463

RESULT 12

Q940V4

AC 0940V4; PRELIMINARY; PRT; 465 AA.

DT 01-DEC-2001 (TREMBlrel. 19, Created)
 DT 01-JUN-2002 (TREMBlrel. 19, last sequence update)
 DE AT330180/T20E20.6.
 OS Arabidopsis thaliana (Mouse-ear cress).
 OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;
 OC Eucosids II; Brassicales; Brassicaceae; Arabidopsis.
 NCBI_taxid=3702;

RA SEQUENCE FROM N.A.
 RA Cheuk R., Chen H., Kim C.J., Koesema E., Meyers M.C., Banh J.,
 RA Bowser L., Carninci P., Dale J.M., Goldsmith A.D., Hayashizaki Y.,

RA Ishida J., Jiang P.X., Jones T., Kamiya A., Karlin-Neumann G.,
 RA Kawat J., Lam B., Lee J.M., Lin J., Liu S.X., Miranda M., Narusaka M.,
 RA Nguyen M., Onodera C.S., Palm C.J., Pham P.K., Quach H.L., Sakurai T.,
 RA Satou M., Seki M., Southwick A., Tang C.C., Toriumi M., Yamada K.,
 RA Yamamura Y., Yu G., Yu S., Shinozaki K., Davis R.W., Theologis A.,
 RA Ecker J.R.;
 RT "Arabidopsis cDNA clones";
 RL Submitted (Aug-2001) to the EMBL/GenBank/DBJ databases.
 RN [2]

RP SEQUENCE FROM N.A.
 RA Kim C.J., Chen H., Cheuk R., Meyers M.C., Banh J., Bowser L.,
 RA Carninci P., Chang E., Dale J.M., Goldsmith A.D., Hayashizaki Y.,
 RA Ishida J., Jones T., Kamiya A., Karlin-Neumann G., Kawat J., Lam B.,
 RA Lee J.M., Lin J., Miranda M., Narusaka M., Nguyen M., Onodera C.S.,
 RA Palm C.J., Quach H.L., Sakurai T., Satou M., Seki M., Southwick A.,
 RA Tang C.C., Toriumi M., Wu H.C., Yamada K., Yamamura Y., Yu G., Yu S.,
 RA Shinozaki K., Davis R.W., Theologis A., Ecker J.R.;
 RT "Arabidopsis ORF clones";
 RL Submitted (Nov-2001) to the EMBL/GenBank/DBJ databases.
 CC -1- SIMILARITY: BELONGS TO THE CYTOCHROME P450 FAMILY.
 DR EMBL: AY052655; AAK96559.1; -
 DR Pfam: PF00067; P450; 1.
 DR InterPro: IPR001128; Cytochrome_P450.
 DR PROSITE: PS00086; CYTOCHROME_P450; UNKNOWN_1.
 DR Heme; Monooxygenase; Oxidoreductase.
 SO SEQUENCE 465 AA; 53814 MW; 280A21D0712FCAL1 CRC64;

Query Match 26.6%; Score 712; DB 10; Length 465;
 Best Local Similarity 32.1%; Pred. No. 4, 3e-43;
 Matches 163; Conservative 94; Mismatches 194; Indels 56; Gaps 8;

QY 14 LPLSLSLILFLIL-----KRNKRTFNLPPGKSGWPLGTTGYLKPATATLGD 67
 DB 3 IMMILGLVITVICTALLRNMQRYSKGLPPTGMPFIFGTEFLKQGP-----DE 57
 QY 68 MOHVSKYKIRSNLFGEPPIVSADAGLNRFILONGRLFECSYPRISIGILGKSMVLV 127
 DB 58 MNORLRYGSFFKSHLIGCPITVSMQAEINRIILNNSGKLVAGYQSMIDLIGTCNIA 117
 QY 128 LVGDHDMRSISLNFSLHARLRTILIKDVERHTLVDSMOONSIFSAODEAKKFTNL 187
 DB 118 VGPSHRLMRGSLSLISPTMKDHLIPKIDPMNMYLCGMDDETVIDQKHKHAFSL 176
 QY 188 MAKHMSMDPEEETEOLKEEYTFMKGYSAPLNLFGLYHKAQSRATILKETERKE 247
 DB 177 SLLDIAETLKKPEVEEYRTEFEKLVGTLSPDIDPGTNYRSGQARNNIDRLTELMO 236
 QY 248 ERKIDKEDEEVEKTEDEAEMSKSDHYRKORTDDLLGWLKHSN-----LSTEQILD 303
 DB 237 ERK-----ESGETF-----DMLGYLMKKEDNRYLLDKERD 269
 QY 304 LILSLFAGHETSSVAIALAIFLOCPKAVEELREHLEIARAKKEGSELNMDYK 363
 DB 270 QVITLISGTEVSTISMAKLYLHDHPALQELREHLEIARAKRPP--DEPLTLDIKS 327
 QY 364 MDTQCVINETLRLGNVRFILRRKALDKVRYKGYDIPSGKVLPIVSAVHLNDSRYD 423
 DB 328 MKFTRAVIYETSLRATIVNGYLKTRDLEINGYLIPKMRIVYTRREINYSLEYDPM 387
 QY 424 LFNWMOONNGASSSGSSTWGNMYMPFGGPRLCAGSELAKLENAVPIHHVLYK 483
 DB 388 IFNFWMEKSLSKS-----YFLFGVGYRLCPGKELGISEVSEFLHYVTYK 436
 QY 484 NWELAEDDQPFAPFVDFPGLPIRVS 510
 DB 437 RWEENGDKLWVFPVFAKGFHLRIS 463

RESULT 13

Q9LH81

AC 09LH81; PRELIMINARY; PRT; 465 AA.

DT 01-OCT-2000 (Tremblrel. 15, Created)
 DT 01-OCT-2000 (Tremblrel. 15, Last sequence update)
 DT 01-MAR-2002 (Tremblrel. 20, Last annotation update)
 DE Cytochrome P450.
 OS Arabidopsis thaliana (Mouse-ear cress).
 CC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 CC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;
 CC eurosids II; Brassicales; Brassicaceae; Arabidopsis.
 NCBI_TaxID=3702;
 RX NCBI_TaxID=3702;
 RP SEQUENCE FROM N.A.
 RC STRAIN-COLUMBIA;
 RA Kaneko T., Kato T., Sato S., Nakamura Y., Asamizu E., Tabata S.;
 RN Submitted (MAR-2000) to the EMBL/GenBank/DBJ databases.
 RL [2]
 RP SEQUENCE FROM N.A.
 RC STRAIN-COLUMBIA;
 RX MEDLINE=20363099; PubMed=10907853;
 RA Nakamura Y.;
 RT "Structural analysis of Arabidopsis thaliana chromosome 3. II.
 RT Sequence features of the regions of 4,251,695 bp covered by ninety pl,
 RT TAC and BAC clones.";
 RL DNA Res. 7:217-221(2000).
 CC -1- SIMILARITY: BELONGS TO THE CYTOCHROME P450 FAMILY.
 DR EMBL: AP002060; BAB02270.1;
 DR InterPro: IPR001128; Cytochrome_P450.
 DR Pfam: PF00067; P450; 1.
 DR PRINTS: PR00385; P450.
 DR PROSITE: PS00086; CYTOCHROME_P450; UNKNOWN_1.
 KW Heme; Monooxygenase; Oxidoreductase.
 SQ SEQUENCE 465 AA; 53862 MW; 630A21D0765EDD0D CRC64;
 Query Match 26.5%; Score 711; DB 10; Length 465;
 Best Local Similarity 32.1%; Pred. No. 5.1e-43;
 Matches 163; Conservative 93; Mismatches 195; Indels 56; Gaps 8;
 QY 14 LLLPSLLSLFLLL-----KRRNRKTRFNLPPGKSGMPFLGRTIGLYKPYATTATLGD 67
 DB 3 IMMMLGLVITVCLCTALRWNNRMRKSKLPPGTMCMPFLGRTTEFLKQGP-----DF 57
 QY 68 MOQVSKYKIKYRNLGEPITVSADAGLNRFILQNERLEECSPRSIGILGKMSLV 127
 DB 58 MKNQRLRYGSEFFKSHILGCPITVGMDELNRYIILNNEKGLVAGYPOGMDLITGICNTAA 117
 QY 128 LVGMHMDMSISLNFSLHARLRTLLKQVRRHTLEVLDSQOONSIFSAODEAKKFTNL 187
 DB 118 VHGSHRLMRGSLSLISPTMMKDLHPKIDFMKNRYLGGWDDLETVDIOEKTKHMAF-L 176
 QY 188 MAKHMSMDPGEETEDOLKKEYVTFMKGVASAPLNLPGTAYHKAQSRATILKFERKME 247
 DB 177 SSLQIAETLKKPEVEYRTFFKLVGTSLVPIDIPGTNTRSGFQANNNIDRLTELMO 236
 QY 248 ERKLIDKEEDOEVEEVTEDEAEMSKSDHVRKORTDDLLGWLVKHSN-----LSTEQILD 303
 DB 237 ERK-----ESGETFT-----DMLGYLMKKEDNRYLLTDKEIRD 269
 QY 304 LILSLFAGHTSSVAIALAIFFLQACPKAVEELREELHETARAKKEGSELMNDYDK 363
 DB 270 QVAVTILSGEYETVSTSMALKYLHDFKALREELRLAIRKRKP--DEPLTDDIKS 327
 QY 364 MDTQCYINETLRGNVVRFLHRAKALKQVRKGYDIPSGWVLPVISAHVHLDNSYDDPN 423
 DB 328 MKFLRAVIFETSRALATVNGVLRKTHDLNGLILPFGWRIYYTRREINIDTSLYEDPM 387
 QY 424 LFNFMWQOONNGASSGSGSFSTWGNMYPFGGPRLCAGSELAKLEMAVFIHHLVKE 483
 DB 388 IFNFMWMEKSLSEKS-----YFLFGGVRRCPEKELGISVSSFLHFTYKY 436
 QY 484 NMELEDDQPPAFVDFPNCPLIRVS 510
 DB 437 RMEENGEDKLAVFPVSPAPKGYHLKCS 463

RESULT 14
 ID Q949P1 PRELIMINARY; PRT; 467 AA.
 AC Q949P1;
 DT 01-DEC-2001 (Tremblrel. 19, Created)
 DT 01-DEC-2001 (Tremblrel. 19, Last sequence update)
 DT 01-JUN-2002 (Tremblrel. 21, Last annotation update)
 DE Putative cytochrome P450 protein.
 GN AT4G19230.
 OS Arabidopsis thaliana (Mouse-ear cress).
 CC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 CC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;
 CC eurosids II; Brassicales; Brassicaceae; Arabidopsis.
 NCBI_TaxID=3702;
 RX NCBI_TaxID=3702;
 RP SEQUENCE FROM N.A.
 RC Goldsmith A.D., Jiang P.X., Lee J.M., Onodera C.S., Quach H.L.,
 RA Tang C., Toriumi M., Yamamura Y., Yu G., Yu S., Bowser L.,
 RA Carlincl P., Chen H., Cheuk R., Hayashizaki Y., Ishida J., Jones T.,
 RA Kamuya A., Karlin-Neumann G., Kawai J., Kim C., Koesema E., Lam B.,
 RA Lin J., Meyers M.C., Miranda M., Narusaka M., Nguyen M., Palm C.J.,
 RA Sakurai T., Satou M., Seki M., Shinn P., Southwick A., Tracy S.E.,
 RA Shinozaki K., Davis R.W., Ecker J.R., Theologis A.;
 RT "Full Length cDNA of gene At4g19230 (GI:7268718)."
 RL Submitted (Aug-2001) to the EMBL/GenBank/DBJ databases.
 RP SEQUENCE FROM N.A.
 RC Yamada K., Banh J., Chan M.M., Chang C.H., Chang E., Dale J.M.,
 RA Deng J.M., Goldsmith A.D., Lee J.M., Onodera C.S., Quach H.L.,
 RA Tang C., Toriumi M., Wu H.C., Yamamura Y., Yu G., Bowser L.,
 RA Carlincl P., Chen H., Cheuk R., Hayashizaki Y., Ishida J., Jones T.,
 RA Kamuya A., Karlin-Neumann G., Kawai J., Kim C., Koesema E., Lam B.,
 RA Lin J., Meyers M.C., Miranda M., Narusaka M., Nguyen M., Palm C.J.,
 RA Sakurai T., Satou M., Seki M., Shinn P., Southwick A., Shinozaki K.,
 RA Davis R.W., Ecker J.R., Theologis A.;
 RT "Arabidopsis Open Reading Frame (ORF) Clones."
 RL Submitted (MAR-2002) to the EMBL/GenBank/DBJ databases.
 CC -1- SIMILARITY: BELONGS TO THE CYTOCHROME P450 FAMILY.
 DR EMBL: AY050980; AAK93657.1;
 DR EMBL: AY091446; AAM14385.1;
 DR InterPro: IPR001128; Cytochrome_P450.
 DR Pfam: PF00067; P450; 1.
 DR PROSITE: PS00086; CYTOCHROME_P450; UNKNOWN_1.
 KW Heme; Monooxygenase; Oxidoreductase.
 SQ SEQUENCE 467 AA; 53037 MW; 2F4230446536D955 CRC64;
 Query Match 25.1%; Score 672.5; DB 10; Length 467;
 Best Local Similarity 32.3%; Pred. No. 3e-40;
 Matches 164; Conservative 96; Mismatches 183; Indels 65; Gaps 15;
 QY 13 LLLPSLLSLFLLLKRRNRKTRFNLPPGKSGMPFLGRTIGLYKPYATTATLGDPMQOHV 72
 DB 10 LFAAGSLFLFLCLISORFFGSKLPLPGTMCMPVYGER---FOLYSQDP-NVFRSQK 65
 QY 73 SKYQKIRSNLGEPTIVSADAGLNRFILQNERLEECSPRSIGILGKMSLVLDGM 132
 DB 66 KRYGSVFKNVLCPCVMISPPAAKFLVTSKHLKPPFPAKBERMLAQALFFHQGD 125
 QY 133 HRDMSISLNFSLHARLRTLLKQVRRHTLEVLDSQOONSIFSAODEAKKFTNLMAKH 192
 DB 126 HAKLKLIVRAIMPESIRN-NVPTDISIADQSLRSR-EGTMINTYDEMKITYTNVALIST 183
 QY 193 MSMDPGEETEDOLKKEYVTFMKGVASAPLNLPGTAYHKAQSRATILKFERKMEER 249
 DB 184 F-----GKDELVLREDLKRCYLLLEKGYNSMPVNLPTFLHKSARKAKELSQLARLSER 239
 QY 250 KLDIKEEDOEVEEVTEDEAEMSKSDHVRKORTDDLLG-VWLKHSNLTSTEQILDILSL 308
 DB 240 R-----QNGSSH-----NDLGSFMDKDEELTDEQIADNIGV 272
 QY 309 LFAGHTSSVAIALAIFFLQACPKAVEELREELHETARAKKEGSELMNDYKKNDDPFG 368

Db 273 IFARDTTASVSMILKYLAEENPNVLAETEOMAI-RKDEGES-LTWGDTKMPILS 330
 QY 369 CVINETRLGNVVFRLHRAKLDVRRKGYDIPSGWVLPYISAVHLNDSRYDQPNLFNFM 428
 Db 331 RVIOETLRVASILSTFFREAVEVEYEGYILPFGWVLPFRNIHNSADIENPGKEDS 390
 QY 429 RMOQNNNGASSSGSFSTWGNMYPFGGPRLCASELAKLEMAVFIHHLVKFNMEL- 487
 Db 391 RE-----VAPKPNTEMPFGNGTHSCPNELAKLEMSIMHILHTTKYSWSTIV 437
 QY 488 -AEDD--OPFAFPVDFPNGLPIRYSR 511
 Db 438 GASDGIQYGFALP-----ONGLPILAR 461

RESULT 15
 Q9FH76 PRELIMINARY: PRT: 463 AA.
 ID 09FH76;
 AC 01-MAR-2001 (TREMblrel. 16, Created)
 DT 01-MAR-2001 (TREMblrel. 16, Last sequence update)
 DT 01-MAR-2002 (TREMblrel. 20, Last annotation update)
 DE Cytochrome P450 (AT5g45340/K9E15_12).
 OS Arabidopsis thaliana (Mouse-ear cress).
 OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;
 OC eucosids II; Brassicales; Brassicaceae; Arabidopsis.
 NCBI_TaxID=3702;
 [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=COLUMBIA;
 RX MEDLINE=20181125; PubMed=10718197;
 RA Sato S., Nakamura Y., Kaneko T., Katoh T., Asamizu E., Kotani H.,
 RA Tabata S.;
 RT "Structural analysis of Arabidopsis thaliana chromosome 5. X. Sequence
 RT features of the regions of 3,076,755 bp covered by sixty pl and TAC
 RT clones.";
 RL DNA Res. 7:31-63(2000).
 RN [2]

RP SEQUENCE FROM N.A.
 RA Cheuk R., Chen H., Kim C.J., Koesema E., Meyers M.C., Banh J.,
 RA Bowser L., Carlucci P., Chang E., Dale J.M., Goldsmith A.D.,
 RA Hayashizaki Y., Ishida J., Jones T., Kamiya A., Karlin-Neumann G.,
 RA Kawai J., Lam B., Lee J.M., Lin J., Miranda M., Narusaka M.,
 RA Nguyen M., Onodera C.S., Palm C.J., Quach H.L., Sakurai T., Satou M.,
 RA Seki M., Southwick A., Tang C.C., Toriumi M., Wu H.C., Yamada K.,
 RA Yamamura Y., Yu G., Yu S., Shinzaki K., Davis R.W., Theologis A.,
 RA Ecker J.R.;
 RT "Arabidopsis cDNA clones."
 RT Submitted (NOV-2001) to the EMBL/GenBank/DBJ databases.
 CC -1- SIMILARITY: BELONGS TO THE CYTOCHROME P450 FAMILY.
 DR EMBL; AB020744; BAB10255.1;
 DR EMBL; AY065065; AAL57698.1;
 DR InterPro: IPR001128; Cytochrome_P450.
 DR Pfam: PF00067; P450; 1.
 DR PRINTS: PR00385; P450; 1.
 DR PROSITE: PS00086; CYTOCHROME_P450; UNKNOWN_1.
 KW Heme; Monooxygenase; Oxidoreductase.
 SQ SEQUENCE 463 AA; 52366 MW; CCD17293F53F812 CRC64;

Query Match 24.5%; Score 658; DB 10; Length 463;
 Best Local Similarity 31.8%; Pred. No. 3.3e-39;
 Matches 163; Conservative 95; Mismatches 184; Indels 70; Gaps 15;

QY 14 LIPSLISLLFLILK-----RRNKKTRFNLPCKSGWPLGFIIGYILKPYTATLGDPM 68
 Db 6 LFLLSAALFLCLRLFLAGVRSSSTKLPLPSTMGPIYGET---FQLTSDDP-NVFF 61
 QY 69 QOHSKYSGKIYRSNLFGEPTIVSADAGLNFETLQNEGRLEFCASYRSIGGILGKMSMLVL 128
 Db 62 AAKRRRGSVKTHVLGCPGVMISSPEAKFVLTKSHLKPFPPEPAKSKRMILGQALFFH 121
 QY 129 VGDHMDRMSISINFLSHARLFTLLKDVHRHTFLVDSWQNSIFSAQDEAKKFTFNLM 188

Db 122 QGDYHSKRLKLVLRAMPDAIN-WPHIESIAOBSLNSWGOTQINTYO-EKQYTFNVA 179
 QY 189 AKHISMDEEET---EOLKREYTFMKGVYSAFLNPGYAVHRALOSRTILKFIK 245
 Db 180 LISIL-----GKDEYVYRDLKRCYIILKGYNSMFINLPGLFHKAMARRELQIILANI 235
 QY 246 MEERLIDKEEDQEEEEVKTDEAMSKSDHYRQRDIDLGLWYK-HSNLSTEQIIDL 304
 Db 236 LSKRR-----QNSSHT-----DLGSPMDKAGLTDQIADN 268
 QY 305 ILSLFAGHETSVAIALAIFPLQACPAVEELREHLEIRAKKEIGESLNDYDKM 364
 Db 269 IIGVIFARDTTASVLMILKYLADNPVLEAVEEOMAIRKDEK-GEK-LTWEDPKM 326
 QY 365 DFQCVINETRLGNVVFRLHRAKLDVRRKGYDIPSGWVLPYISAVHLNDSRYDQPNL 424
 Db 327 PLTVRYIOETLRVAATILSTFFREAVEVEYEGYILPFGWVLPFRNIHNSADIENPGK 386
 QY 425 FNPWRMOQNNNGASSSGSFSTWGNMYPFGGPRLCASELAKLEMAVFIHHLVKFN 484
 Db 387 FDPSEFE-----VAPKPNTEMPFGSGTHSCPNELAKLEISVLIHILHTTKYR 433
 QY 485 WELAEDDQ-----PFAPFPVDFPNGLPIRYSR 511
 Db 434 WSIQPSDGIQYGFALP-----ONGLPILAR 461

Search completed: June 10, 2003, 21:58:53
 Job time : 52 secs

